Investigating the Evolutionary Origins and Modification of Novel Morphologies and their Developmental Networks

by

William J. Glassford

B.S., Indiana University of Pennsylvania

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

2016

UNIVERSITY OF PITTSBURGH

Dietrich School of Arts and Sciences

This dissertation was presented

by

William J. Glassford

It was defended on

May 31st, 2016

and approved by

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

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

Dr. Jeffrey D. Hildebrand, Associate Professor, Department of Biological Sciences

Dr. Veronica F. Hinman, Associate Professor, Department of Biological Sciences,

Carnegie Mellon University

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

Investigating the Evolutionary Origins and Modification of Novel Morphologies and their Developmental Networks

William J. Glassford

University of Pittsburgh, 2016

INVESTIGATING THE EVOLUTIONARY ORIGINS AND MODIFICATION OF NOVEL MORPHOLOGIES AND THEIR DEVELOPMENTAL NETWORKS

William J. Glassford, PhD

University of Pittsburgh, 2016

The nature of the origin of morphological characters has long been a central subject of interest in the field of evolutionary developmental biology. Currently, many morphologies are known to be underscored by vast gene regulatory networks (GRNs) such that GRNs are anticipated for any feature of anatomy. Hence, if networks drive anatomical development, how do they evolve? The co-option of networks, a phenomenon in which cohorts of pre-existing transcriptional circuits are redeployed to new developmental settings, has been proposed to facilitate the rapid evolution of GRNs. Although several examples suggest the contribution of network co-option to the evolution of novel structures, examples that demonstrate and explore this process in molecular detail are currently lacking. In this dissertation I investigate the posterior lobe as a model of network co-option. A cuticular outgrowth on the genitalia of male fruit flies, this morphology is unique to the *Drosophila melanogaster* clade. By studying the ancestry of one gene's posterior lobe activity, I discovered that it existed before the evolution of the posterior lobe, and had been redeployed from a network active during embryonic life. I next investigate the origin of the posterior lobe by studying the intercellular signaling pathways that contribute to its specification, discovering that a drastically altered pattern of the Notch ligand *Delta* is necessary for the development and evolution of the posterior lobe. I then explore how an embryonic circuit that was co-opted to the posterior lobe was subsequently modified to alter its shape. Finally, I study the origins of novelty

at the level of an individual transcriptional circuit, analyzing all possible intermediate states along its evolutionary path. These studies demonstrate the value of an approach focused on understanding the co-option and origination of regulatory circuitry for the study of the evolution of novel characters.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.0 INTRODUCTION

Morphological novelties, newly acquired anatomical structures such as wings, flowers, or the neural crest, are charismatic examples of the vast diversity in form that exist in the kingdoms of life. Underlying the development of complex morphologies are intricate webs of gene regulatory connections that orchestrate thousands of gene expression decisions to drive the cell growth, migration and cell shape changes that underlie the formation of anatomy. The nature of how such complex morphologies and their associated gene regulatory networks (GRNs) are generated is a major topic of interest to those studying the evolution of development. One notion to explain the construction of these complex GRNs is that new structures may re-use parts of networks that were already in place. This model of "network co-option" proposes that the ectopic expression of regulatory genes that reside near the top of a GRN hierarchy can initiate the redeployment of lower-tier genes into the novel location through their pre-existing circuits (Gompel, Prud'homme, Wittkopp, Kassner, & Carroll, 2005; Monteiro & Podlaha, 2009). Hence, the co-option model allows for a complex GRN to evolve in relatively few steps. The evolution of the beetle horn (Moczek, Rose, Sewell, & Kesselring, 2006; Moczek & Nagy, 2005), the sea urchin larval skeleton (Gao & Davidson, 2008) and the butterfly wing's eye spot (Keys et al., 1999) are all examples of novelties for which a co-option mechanism has been proposed. In each one of these cases, a set of genes expressed in these tissues is known to participate in other networks that are clearly ancestral, existing before these novelties arose. However these cases are

technically limited to inferences based upon the description of gene expression patterns, and further cannot address the causative changes underlying their origination. Therefore, to understand how GRNs evolve to generate novelties, a system in which the direct analysis of the regulatory circuitry responsible for redeploying genes into a novel GRN is required.

In this thesis, I will investigate an example of a recently evolved trait to answer three main topics of interest: 1: how it co-opted pre-existing regulatory DNA in the evolution of its developmental network, 2: identification of upstream genes that may have initiated the co-option of this GRN, and 3: how its components became individualized from their ancestral activities to contribute to the diversification of this novel structure.

1.1 THE COMPOSITION AND EVOLUTION OF GENE REGULATORY NETWORKS

1.1.1 Metazoans Utilize a Shared "Toolkit" of Developmental Genes

One of the most surprising findings of the post-genomic era was the discovery that metazoan (multicellular) species generally lack uniqueness in their repertoire of developmental patterning genes despite sweeping differences in their complexity and organization. One proxy measure for complexity is the number of different cell types an organism possesses, as they represent the basic building blocks of bodies and embryos (Valentine, 2003). Sponges are some of the simplest metazoans, with as few as 4 cell types in some species (Simpson, 1984), while hundreds of cell types have been identified in humans (Alberts et al., 1989). However, this stark contrast in complexity is not mirrored by differences in the relative number of genes: the human genome

may have as few as ~19,000 protein-coding genes in its genome (Ezkurdia et al., 2014), whereas the genome of the sponge *Amphimedon queenslandica* was found to have at least 18,693 protein coding loci with identifiable homologs (Srivastava et al., 2010).

Within metazoan genomes, pioneering investigations of the developmental patterning gene repertoire uncovered the discovery that species across metazoa exhibit a shared "toolkit" of transcription factors and signaling pathways (Carroll, Grenier, & Weatherbee, 2004). One of the most important processes to a multicellular organism is the ability to specialize cells into distinct types. This coordination is regulated by intracellular signaling pathways that are shared between many species (Nichols, Dirks, Pearse, & King, 2006), and recent genome sequencing has revealed that the ligands and effectors of several of these pathways have existed since the ancestor of all extant metazoans (Srivastava et al., 2010). Another developmental process that is deeply conserved is the specification of the animal body plan along the anterior/posterior axis. The identity of segments along this axis is governed by Hox transcription factors that are clustered in gene complexes conserved between vertebrates and invertebrates (Duboule, 2007). In addition, many other transcription factors have maintained their developmental roles and specificities (Nitta et al., 2015). For example, both the *Drosophila* gene *eyeless* and its vertebrate homologue *Pax6* can induce ectopic eyes when ectopically expressed during development in *Drosophila* (Halder, Callaerts, & Gehring, 1995). Additionally, the vertebrate and invertebrate homologues of achaete-scute maintain their pattern and function in neural development (Johnson, Birren, & Anderson, 1990). The deep preservation of gene function observed in many signaling pathways and transcription factors suggests that their protein coding regions are not a major source of genetic variation contributing to animal diversity.

1.1.2 Enhancers Regulate the Activity of Genes in Time and Space

An increasing number of examples has indicated that much of morphological evolution is driven by changes to *cis*-regulatory elements known as enhancers that control gene transcription (Carroll, 2008; Stern & Orgogozo, 2008; Wray, 2007). Therefore, an understanding of the relationship between enhancers and the genes they regulate is necessary to understand the evolution of animal morphology. A schematic of a typical gene, and the location of its enhancer elements is presented in Figure 1.1A. Enhancers may be located within the immediate surrounding non-coding DNA and intronic non-coding DNA (Figure 1.1A, center and right gray boxes), but can also regulate the activity of genes from a very distant genomic location (left gray box) (Symmons & Spitz, 2013). Enhancers are typically between 200 and 1000 base pairs in length, and and consist of multiple transcription factor binding sites (Figure 1.1B) (Levine, 2010). Transcription factors utilize several distinct mechanisms to regulate transcription, including the recruitment of chromatin modifying cofactors to alter chromatin accessibility and the recruitment of the transcriptional initiation factors including RNA polymerase II to the promoter (Kadonaga, 2004).

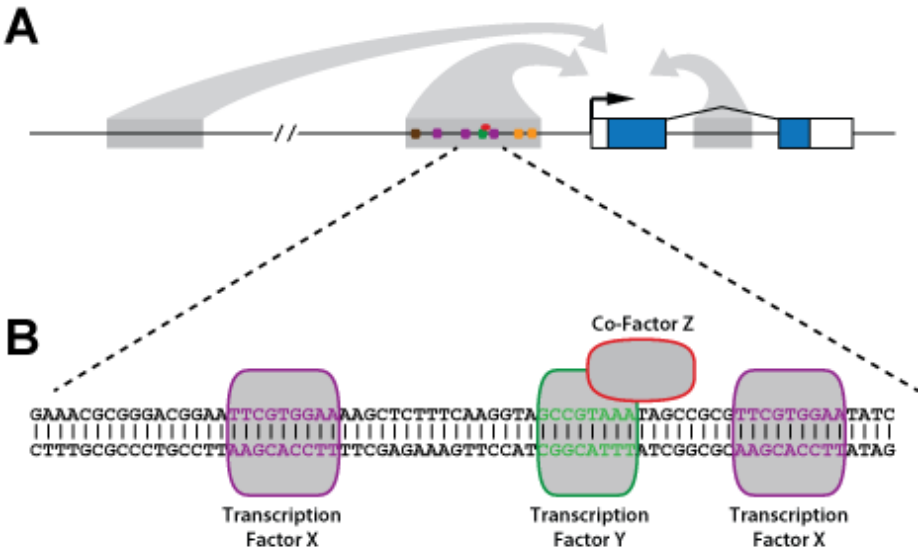

Figure 1.1. **Gene expression is regulated by *cis*-regulatory elements called enhancers.**

(A) Schematic of a model gene and its regulatory region. Enhancers (gray) are located in non-coding DNA surrounding the exonic portion of a gene (blue and white boxes), and can be very distant from the promoter. (blue boxes) coding exonic sequence. (white boxes) 5' and 3' non-coding exon regions. (// symbol) denotes a long distance between promoter and enhancer that was truncated for the purpose of presentation. (B) Enhancers are comprised of multiple bindings sites (colored text) recognized by transcription factors (TFs X and Y). Transcription factors regulate transcription through the recruitment of chromatin modifying cofactors (Co-factor Z) to alter chromatin accessibility and transcription initiation factors to the promoter (black arrow).

Individual enhancers can regulate separate patterns (Dynan, 1989; Levine, 2010) (Figure 1.2), or similar patterns (Hong, Hendrix, & Levine, 2008) of the given gene's expression, conferring a great deal of modularity or independence to individual portions of the regulatory DNA. Two or more enhancers that regulate the same pattern of expression have been described as "shadow" enhancers, which may help maintain robust expression in the face of environmental or genetic perturbations (Frankel et al., 2010; Swami, 2010) and may afford an extra degree of flexibility to modify a gene's enhancers without disruption of critical functions (Hong et al.,

2008). Due to this modularity, evolutionary modifications to enhancer DNA is predicted to be less pleiotropic than changes to proteins (Carroll, 2008; Stern & Orgogozo, 2008; Wray, 2007). For example, the model gene presented in Figure 1.2 is expressed in three distinct tissues during development. Modification of any individual enhancer would likely only alter one activity, whereas mutations in the coding region of the gene may incur pleiotropic effects in multiple, if not all activities (Wray, 2003).

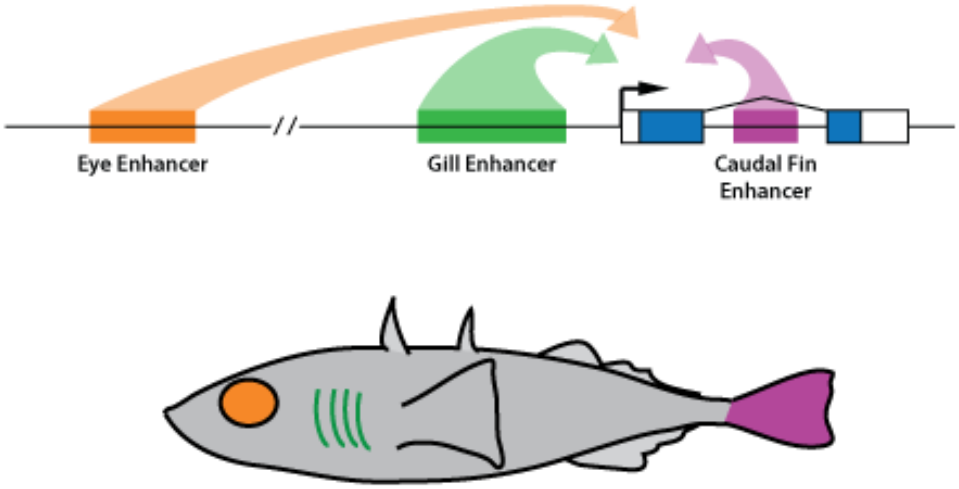

Figure 1.2. *Cis*-regulatory elements (CREs) are often modular in nature.

Schematic of a model gene and its regulatory region. Several CREs (orange, green and purple bars) regulate a model gene's (blue and white boxes) expression in three distinct expression patterns. Colors of enhancers correspond to their pattern of expression (bottom). // symbol denotes a long distance between promoter and CRE that was truncated from the schematic for the purpose of presentation.

1.1.3 Mechanisms of Evolution of Novel Enhancer Activities

In the pursuit of a GRN's origin, a critical question to consider is how new enhancers arise in the first place. Four major mechanisms have been proposed to account for the evolution of novel enhancer activities: transposition, co-option, promoter switching, and *de novo* (Rebeiz, Patel, & Hinman, 2015).

Transposition

A long-appreciated mechanism for the evolution of novel enhancer activities is the insertion of a transposable element into the regulatory region of a gene (Britten, 1996). If the inserted element contains a combination of binding sites sufficient to drive expression, it could confer a new expression pattern to the gene through a single mutational event (Figure 1.3A). Historically, many functional regulatory sequences have been shown to contain sequences with homology to transposons, and yet few of these have been tested for their sufficiency to drive expression (Britten, 1996). Although not a significant crop pest, *Drosophila melanogaster* evolved resistance to DDT by increasing the levels of cytochrome p450 gene *Cyp6G1* in the gut (Daborn, Boundy, Yen, Pittendrigh, & Ffrench-Constant, 2001). This was later demonstrated to be due to an insertion of the *Accord* transposon into the promoter region of *Cyp6G1*, which increased gut expression levels and slightly altered its deployment pattern (Chung et al., 2007). Widespread insertion of transposons can rapidly build a novel GRN responsive to a common set of trans-activators, and has been implicated in the evolution of the immune system (Chuong, Elde, & Feschotte, 2016), and during the evolution of pregnancy (Lynch, Leclerc, May, & Wagner, 2011).

Promoter Switching

Gene promoters are known to have several mechanisms that police the specificity of an enhancer to a single promoter (Gaszner & Felsenfeld, 2006), representing mechanisms that can be modified to change an enhancer's target promoter. One mechanism that could cause such "promoter switching" is a chromosomal rearrangement such as an inversion or a deletion, which can alter the relative position of enhancers, promoters and the elements that regulate their interaction (Cande, Chopra, & Levine, 2009; Levine, 2010), (Figure 1.3B). An additional mechanism to induce promoter switching would be a direct alteration to the tertiary structure that organizes the chromosome into topologically associated domains (TADs). CTCF is a structural protein that directly binds to DNA and acts as an anchor for proteins such as cohesin to organize the chromosome into domains (Ong & Corces, 2014). An analysis of evolving TADs revealed that they correlated with the gain and loss of CTCF binding sites and with the presence of cohesin (Vietri Rudan et al., 2015). In addition, a comparison of closely related *Drosophila* species found evidence for positive selection in regions containing novel CTCF binding sites and that gain and loss of CTCF binding correlated with diverging patterns of neighboring gene expression (Ni et al., 2012). Inversions and duplications can also disrupt TADs, such as those implicated in cases of polydactyly (Lupianez et al., 2015).

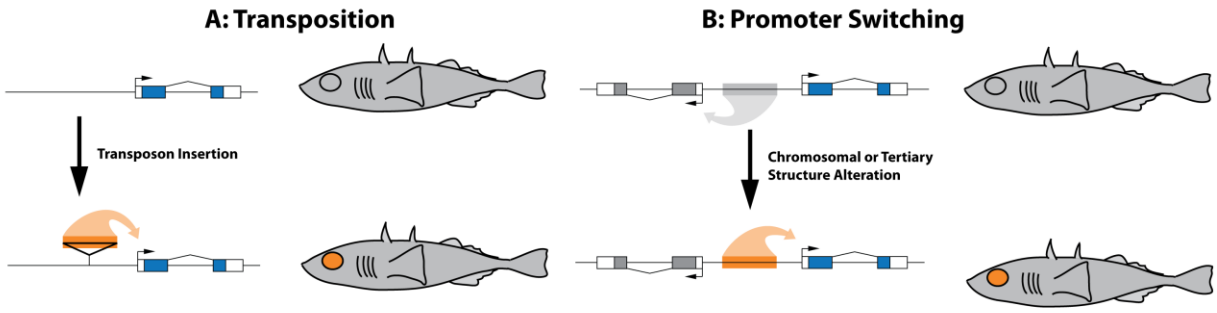

Figure 1.3. Transposition and Promoter Switching Mechanisms of Evolution of Novel Enhancer Activities.

(A) A transposon containing a functional enhancer is inserted near a gene inducing a novel expression pattern. (B) A chromosomal alteration (for example an inversion or deletion) or a change to the tertiary structure of the chromosome may redirect a pre-existing enhancer to a new gene to induce a novel pattern of expression.

Co-option

While promoter switching would cause a new promoter to adopt a pre-existing enhancer to drive a pre-existing pattern of expression, enhancers themselves may be modified to drive additional patterns of expression. As mentioned earlier, enhancers are composed of multiple transcription factor binding sites and drive specific patterns of expression by virtue of which combinations of factors exist in a given tissue (Figure 1.1B) (Davidson, 2006). For a pre-existing enhancer to be co-opted to drive a novel enhancer activity, the addition (or subtraction) of binding sites to the element would combine with pre-existing sites to create a combination of factors sufficient to drive expression in a new tissue (Rebeiz, Jikomes, Kassner, & Carroll, 2011) (Figure 1.4A). This mechanism contributed to the evolution of several novel pigmentation patterns that adorn the body of *Drosophila guttifera*. Novel patterns of the Wnt signaling ligand *wingless* correlate with novel abdomen, thoracic and wing pigmentation patterns (Werner,

Koshikawa, Williams, & Carroll, 2010). An analysis of the regulatory regions responsible for inducing novel expression revealed that they overlapped with ancestral eye and antennal imaginal wing disc activities of *wingless* (Koshikawa et al., 2015), suggesting that the new patterns evolved by co-opting binding sites contained within these ancestral enhancers.

De novo evolution of enhancer elements

Finally, an enhancer may evolve *de novo* from non-functional precursor DNA. In this case, the novel enhancer is composed of binding sites that had no regulatory function prior to the evolution of the new enhancer (Figure 1.4B). *De novo* evolution is essentially the null hypothesis to consider when testing for co-option evolution. A screen for novel enhancers was performed on pseudogenized exons in Zebrafish, as exons presumably do not contain regulatory information (Eichenlaub & Ettwiller, 2011). Several novel enhancers were identified (Eichenlaub & Ettwiller, 2011), suggesting that pre-existing components are not necessary to create novel enhancers. One caveat to this analysis is that the whole genome duplication that permitted mass pseudogenization in zebrafish occurred 350 million years ago (Eichenlaub & Ettwiller, 2011). Given this long span of time, it remains experimentally difficult to prove that its ancestral regulatory DNA had no activity.

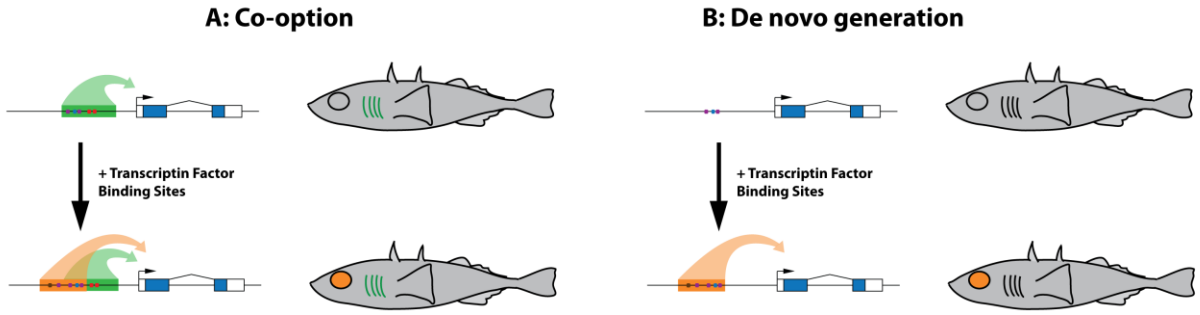

Figure 1.4. Co-option and *de novo* Mechanisms for the Evolution of Novel Enhancer Activities.

(A) Novel transcription factor binding sites combine with pre-existing transcription factor binding to initiate a novel enhancer activity (B). During the *de novo* origination of an enhancer, novel transcription factor binding sites are built up to create a novel activity in a region that had no prior regulatory activity.

1.1.4 The Network Co-option Mechanism

Alterations to individual gene activities are generally considered to be insufficient to induce the formation of complex structures. Therefore, in order to evolve a complex novel morphology a large number of novel gene activities must be organized to construct a novel GRN. One mechanism proposed to contribute large numbers of genes to a novel GRN is network co-option (Gao & Davidson, 2008; Moczek & Nagy, 2005) (depicted in Figure 1.5). In a network co-option event, downstream genes in an ancestral GRN are expressed in a novel tissue through the reutilization of their pre-existing enhancers (Figure 1.5C). These enhancers are active in the new location due to a novel activity of their upstream regulator, thus recapitulating the ancestral network topology in the novel GRN. This mechanism of GRN evolution is predicted to require fewer modifications than building a novel GRN through a novel enhancer for every gene activity

(Monteiro & Podlaha, 2009) (Figure 1.5D). Below, I will discuss examples that illustrate the co-option mechanism, and explore gaps in our current understanding of GRN origination.

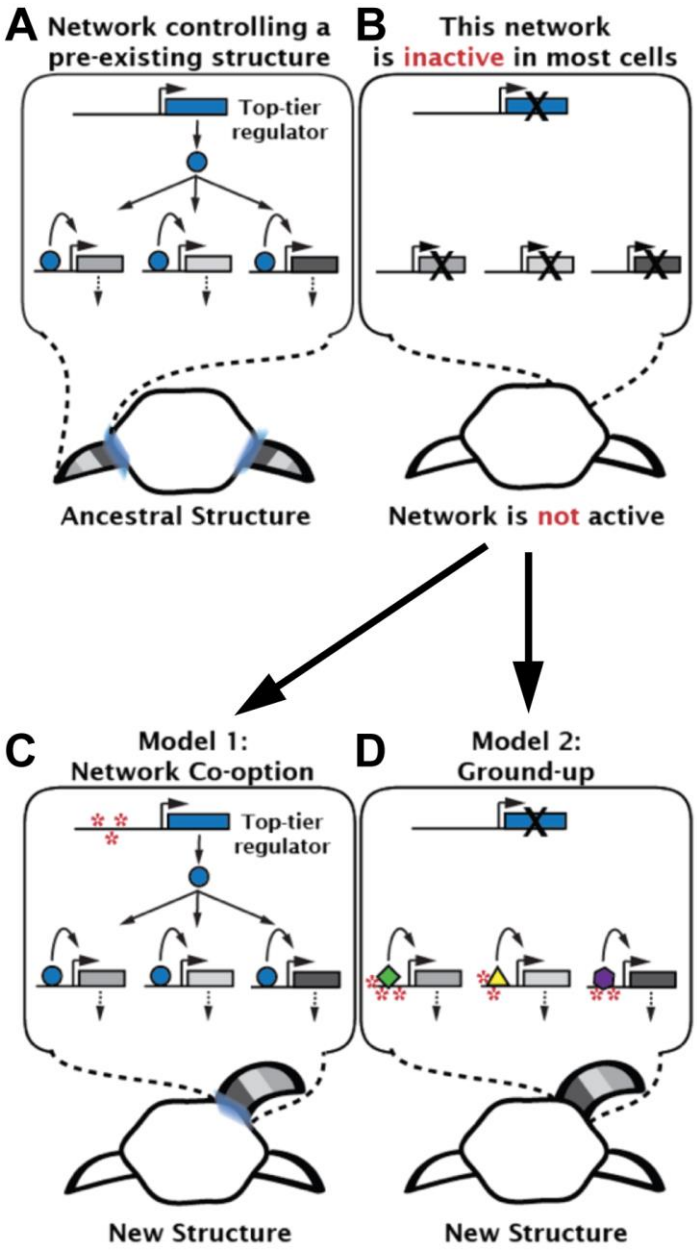

Figure 1.5. Models Depicting the Network Co-option Mechanism.

(A) An ancestral gene regulatory network (GRN) consists of a top tier transcription factor (blue) that regulates the activity of three downstream genes (grey) to govern the development of an ancestral structure. (B) This network is inactive in most other tissues. (C) In a network co-option model, a novel activity of the top-tier regulator

redeploys the downstream components of the ancestral GRN using the pre-existing circuit. (D) In an alternative "ground-up" model every component of the novel GRN requires a novel enhancer and no pre-existing circuitry is co-opted. The construction of the novel GRN by either mechanism leads to the evolution of a novel morphology (bottom), but the network co-option mechanism requires fewer steps (red asterisks).

1.2 EXAMPLES OF NETWORK CO-OPTION AND DIVERSIFICATION

1.2.1 The Origin and Diversification of the Beetle Horn

The beetle horn is a morphological novelty present in thousands of beetle species in the scarab superfamily and is used as a weapon during competition for resources (Arrow, 1951; Emlen, Lavine, & Ewen-Campen, 2007). Two networks have been proposed to have been co-opted during the evolution of the beetle horn: the limb patterning network and the insulin signaling pathway (Emlen, Szafran, Corley, & Dworkin, 2006). Several transcription factors of the insect appendage network are expressed during the formation of the horn pronotum, and are thought to establish a proximal-distal axis in a manner similar to ancestral appendages (Moczek & Nagy, 2005). The beetle horn is polyphenic, or sensitive to environmental factors (Moczek & Nagy, 2005); members of the insulin growth factor (IGF) pathway are sensitive to available nutrients and have been shown to contribute to beetle horn development and nutrient sensitivity (Emlen, Warren, Johns, Dworkin, & Lavine, 2012). The beetle horn is also sexually dimorphic: the males in horned beetle species display large horn ornaments, but their female counterparts often exhibit no horn or rudimentary horns (Arrow, 1951). Differential splicing of the transcription factor

doublesex (*dsx*) controls somatic cell sexual differentiation in insects (Burtis & Baker, 1989), and has been found to be expressed during beetle horn development (Snell-Rood et al., 2011).

Several of the genes that contribute to the development of the beetle horn have subsequently diverged in activity between horned species. *dsx* controls sex and morph-specific horn development in the horned beetle *Onthophagus taurus*, but its function and pattern of expression has been altered in its close relative *O. sagittarius* (Kijimoto, Moczek, & Andrews, 2012). In addition to their role in the origination of the beetle horn, the appendage network genes *Distalless* (*Dll*) and *homothorax* (*hth*), have evolved differential activity between *O. taurus* and closely related species *O. binodis* (Moczek & Rose, 2009). These findings suggest that the networks for beetle horns have been quite plastic since their co-option from the appendage network.

1.2.2 The Origin and Diversification of the Sea Urchin Larval Skeleton

There are several morphological differences between larvae of the phylum *Echinodermata*, including the presence of a larval skeleton in sea urchins that is not present in larvae of the more basal class containing sea stars (Hyman, 1955). The larval skeleton is secreted by a derived lineage of the embryonic mesoderm called the skeletogenic mesoderm (SM) (Gao & Davidson, 2008). The SM shares a large number of genes with juvenile skeletogenic centers, which contribute to the adult skeleton, and it is thought that the SM co-opted circuits from the juvenile tissue during its origination (Gao & Davidson, 2008). The sea cucumber also exhibits a mesoderm-derived larval skeleton that expresses many of the same network components as the sea urchin GRN (McCauley, Wright, Exner, Kitazawa, & Hinman, 2012), suggesting that the SM

was co-opted in a common ancestor. Differences in the complexity of larval skeleton morphology and in the content of the sea cucumber and sea urchin GRNs indicate that the skeletogenic network has since been modified in the intervening time of divergence (McCauley et al., 2012).

1.2.3 The Origin and Diversification of the Butterfly Wing Spot

Butterfly eyespots are circular bullseye marks formed through the patterning of brightly colored scales on the wings of butterflies of the Nymphalidae family (Stevens, 2005), and have been shown to provide a role in predator avoidance (Prudic, Stoehr, Wasik, & Monteiro, 2015). Several ancestral GRNs have been advanced as the ancestral network that was co-opted during the origin of the eyespot GRN, including the network responsible for patterning the anterior/posterior axis of insect body segments (Keys et al., 1999) and the wound healing network (Monteiro, Glaser, Stockslager, Glansdorp, & Ramos, 2006). Four of five eyespot genes assayed for activity in a survey of the Nymphalids were found to have originated concurrently, suggesting that the eyespot GRN likely evolved in a single origination event (Oliver, Tong, Gall, Piel, & Monteiro, 2012). Several of the genes had been lost during the nymphalid radiation, possibly indicative of network "simplification," or a pruning of unnecessary components brought to spot development, while still maintaining the genes necessary to induce the phenotype (Oliver et al., 2012).

1.2.4 The Role of Regulatory DNA in Network Co-option, Diversification and Origination

While network co-option events have been inferred through the identification of gene expression profiles that resemble other pre-existing networks, this phenomenon has surprisingly not been characterized at the level of a network's regulatory circuits. Regulatory DNA is central to the network co-option hypothesis as it is the vehicle through which co-opted gene activities are redeployed. It has been proposed that network co-option events could be traced by exploring how activities of CREs correspond to one another (Monteiro & Podlaha, 2009). Showing that the ancestral state of the enhancer is capable of being redeployed into a novel location to prove the efficacy of any network co-option hypothesis.

Studying co-opted enhancers is important for addressing questions about the diversification of co-opted networks as well. In all three examples of network co-option presented in the previous section, the networks were subsequently modified between the species that develop these novelties. How were these networks able to diversify without pleiotropic consequences? One potential mechanism that would reduce the pleiotropic link between novel and ancestral networks is the individualization of enhancers at co-opted loci, allowing them to be modified without pleiotropic constraint (Figure 1.6). Did the enhancers separate after co-option; do they need to? Identifying co-opted enhancers would allow one to study how a network has been individualized following a co-option event.

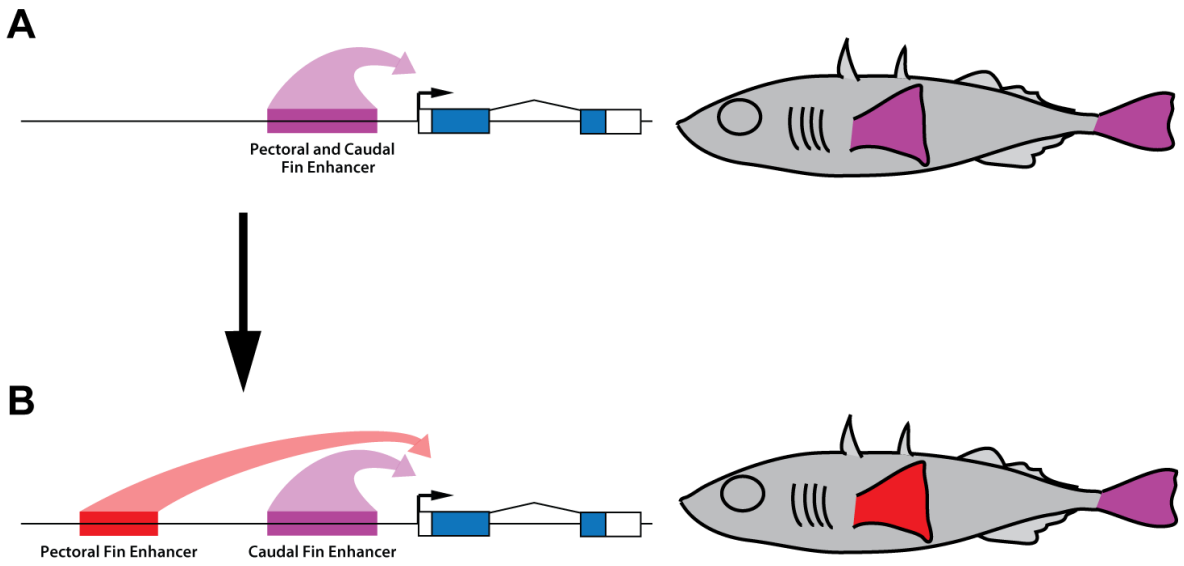

Figure 1.6. Model depicting the individualization of a pleiotropic enhancer.

(A) Schematic of a gene and an enhancer that drives expression in two tissues. (B) The evolution of a novel enhancer for one activity and the loss of the capability of the ancestral enhancer to drive expression in that activity separates the two enhancers into distinct regulatory regions, potentially reducing pleiotropy.

Studying the regulatory DNA governing the activity of top tier genes in GRNs that control novel structures is likely necessary to understand the molecular mechanisms underlying their origination. Unlike the lower tiers of a co-opted GRN that are predicted to have multifunctional enhancers, top-tier regulators are predicted to have independently gained novel gene activities (Stern & Orgogozo, 2008). It can be difficult to trace evolutionary changes to regulatory DNA as it exhibits a rapid turnover of transcription factor binding sites (Ludwig, Patel, & Kreitman, 1998; Swanson, Schwimmer, & Barolo, 2011). Studying the evolution of regulatory DNA may therefore be difficult for the three examples listed in the previous section as many of them originated deep in the past. The basal nymphalid subfamily without eyespots diverged 90 million years ago (Oliver et al., 2012). The larval skeleton in echinoderms might be

quite old: the brittle star class, which is basal to the sea urchin and sea cucumber classes, also exhibits a larval skeleton (McCauley et al., 2012), although it has not been assayed for the skeletogenic network. Finally, the several thousand beetle species that exhibit the beetle horn morphology are spread throughout the scarab beetle superfamily (Arrow, 1951; Emlen et al., 2007), which is over 100 million years old (Ahrens, Schwarzer, & Vogler, 2014). A more recently derived morphology may be more amenable for an analysis of the evolution of regulatory DNA in a network co-option event.

1.3 DROSOPHILID EXTERNAL MALE GENITALIA AS A MODEL SYSTEM

1.3.1 Male Genitalia Exhibit Rapidly Evolving Morphologies

External male genitalia represent some of the most rapidly evolving morphologies among animals (Eberhard, 1985). Insects contain some of the most speciose orders in the animal kingdom, and are estimated to include 2.6-7.8 million species (Stork, McBroom, Gely, & Hamilton, 2015). This species diversity is often accompanied by a corresponding diversity in external male genitalia, which are often used to discriminate different insect species (Eberhard, 1985). Perhaps the most well studied insect is the model species *Drosophila melanogaster*. *D. melanogaster* and its close relatives exhibit many diverse sexually dimorphic morphologies (Kopp & True, 2002), including a novel structure unique to the *D. melanogaster* clade called the posterior lobe.

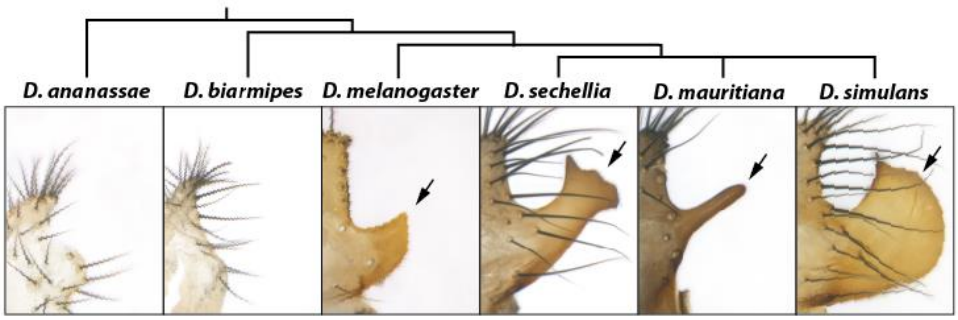

Figure 1.7. The posterior lobe is a novel morphology unique to the posterior lobe clade.

Tree depicting the phylogenetic relationships of selected *Drosophila* species, and images of their lateral plate morphologies. The posterior lobe is an outgrowth of the lateral plate (arrows) that is unique to lobed species.

The posterior lobe is a cuticular outgrowth on the male genitalia of all four *D. melanogaster* clade species (Figure 1.7, arrows). During copulation the posterior lobe inserts between the female seventh and eighth abdominal segments (Kamimura, 2010) and is necessary for genital coupling (Frazee & Masly, 2015), but does not directly contact the female genitalia (Kamimura, 2010). The *D. melanogaster* clade last shared a common ancestor 300,000-900,000 years ago (Tamura, Subramanian, & Kumar, 2004) and their lobes have since diverged in shape and size.

1.3.2 The Posterior Lobe as a Model of Gene Regulatory Network Evolution

As a recently derived morphology in a widely-studied model organism, the posterior lobe presents a great opportunity to study the evolution of a novel GRN at the level of its constituent enhancers. In this thesis, I identify and analyze multiple posterior lobe network enhancers to study the co-option, origination and individualization of the posterior lobe's developmental network.

2.0 CO-OPTION OF AN ANCESTRAL HOX-REGULATED NETWORK UNDERLIES THE EVOLUTION OF THE POSTERIOR LOBE

This chapter was published in *Developmental Cell* Volume 34, Issue 5, on the 14th of September 2015, Pages 520–531. The following collaborators contributed data presented in this chapter: Mark Rebeiz helped design and write the paper, and characterize *upd* expression. Sarah Smith screened the *upd* locus for enhancers. Yang Liu and Natalie Dall helped to characterize *eya* activity, and Yang screened the *eya* locus for enhancers. Werner Boll and Markus Noll analyzed the *Poxn* mutant phenotype in the posterior spiracle.

2.1 INTRODUCTION

"structural genes are building stones which can be used over again for achieving different styles of architecture…evolution is mostly the reutilization of essentially constituted genomes"

-Emile Zuckerkandl, 1976 (Zuckerkandl, 1976)

Evolutionary biologists have long been intrigued by the origins of biological complexity. While the complexity of living systems can be considered at multiple levels of organization (e.g. the origins of DNA-based life (Crick, 1968; Orgel, 1968), organelles (Sagan, 1967), or multicellularity (Bonner, 1998)), the evolutionary origin of morphological complexity is a

developmental problem (Muller & Wagner, 1991). Morphological structures are patterned and formed during the process of embryonic development, and each cell in the developing organism must derive unique physical properties from an identical DNA code. This apparent paradox is solved by differential gene activity, governed by vast gene regulatory networks (GRNs) (Davidson, 2001). Regulatory factors within GRNs bind transcriptional regulatory sequences such as enhancers to combinatorially determine the expression status of each gene of the network in morphological space and developmental time (Small, Blair, & Levine, 1992). Hence, an understanding of the origins of morphological complexity necessitates investigations into how GRNs originate.

A growing body of evidence has implicated the re-use, or co-option, of existing networks in the evolution of novel morphological structures (Gao & Davidson, 2008; Keys et al., 1999; Kuraku, Usuda, & Kuratani, 2005; Moczek & Nagy, 2005). For example, expression of the appendage-patterning network within the developing beetle horn suggests that this novelty arose through the establishment of a new proximo-distal axis (Moczek et al., 2006; Moczek & Nagy, 2005; Moczek & Rose, 2009). Such findings evoke a scenario in which a cohort of downstream appendage enhancers were in turn activated in the new setting, generating a unique developmental output. However, instances of co-option have traditionally been supported by correlations in gene expression, relationships that may arise without the reuse of existing circuits (Abouheif, 1999). Currently, examples that illustrate this phenomenon at the level of enhancers and the constituent binding sites that were co-opted are lacking.

Here, we trace the evolutionary history of the posterior lobe, a recently evolved morphological structure present in the model organism *Drosophila* (*D.*) *melanogaster* at the level of its network, enhancers, and the transcription factor binding sites of which these are composed.

2.2 RESULTS

2.2.1 The Posterior Lobe is a Morphological Novelty Unique to the *D. melanogaster* Subgroup

Male genitalia represent the most rapidly evolving morphological structures in the animal kingdom (Eberhard, 1985), and are often used to taxonomically distinguish insect species. The posterior lobe is a hook-shaped outgrowth unique to the external genitalia of *D. melanogaster* and its closest relatives in the *melanogaster* clade (Figure 2.1) (Jagadeeshan & Singh, 2006; Kopp & True, 2002). A cuticular projection similar to the posterior lobe is also present in the *yakuba* clade (Yassin & Orgogozo, 2013), suggesting a recent origin of this structure in the *melanogaster* subgroup (Figure 2.2). Among members of the *melanogaster* clade, the posterior lobe is highly divergent in shape and size, and represents the only reliable character to distinguish species identity (J. Coyne, 1993). During mating, the posterior lobe is used by the male to grasp the female ovipositor (Jagadeeshan & Singh, 2006), and subsequently is inserted between cuticular plates at the posterior of the female abdomen during genital coupling (Robertson, 1988). Given the recent evolution of the posterior lobe, and its presence in *D. melanogaster*, a highly tractable model organism for studying the structure and evolution of gene regulatory networks, we sought to elucidate its evolutionary origins.

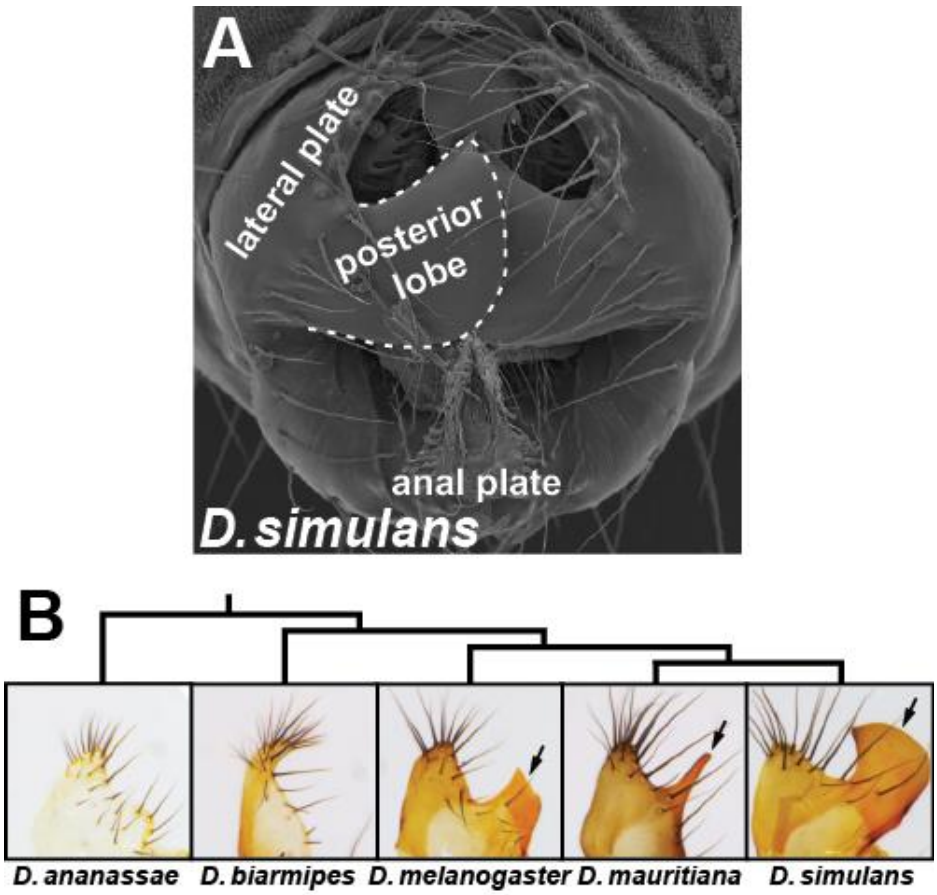

Figure 2.1. The Posterior Lobe is a Morphological Novelty Unique to the *D. melanogaster* Clade.

(A) Scanning electron micrograph of a *D. simulans* male with important structures labeled. (B) Tree depicting the phylogenetic relationships of the species in this study, and brightfield images of their lateral plate cuticle morphologies. The posterior lobe is an outgrowth of the lateral plate unique to the *melanogaster* clade (arrows).

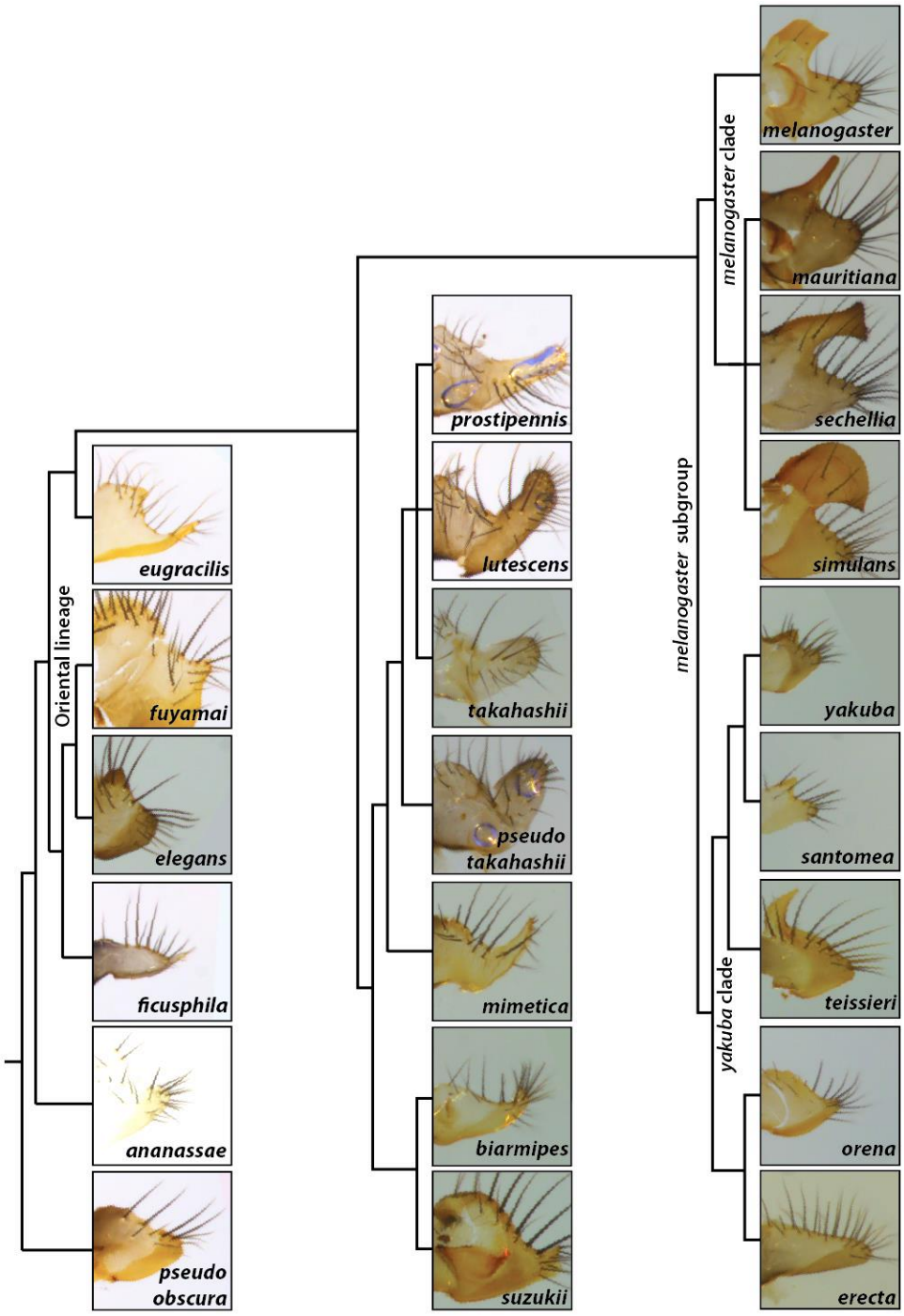

Figure 2.2. A Survey of Lateral Plate Morphology and Novelty in the Oriental Lineage.

Tree depicting the phylogenetic relationships of selected species in the oriental lineage of *Drosophila*, as well as outgroup species in the *montium*, *ananassae*, and *obscura* groups. The posterior lobe is a cuticular outgrowth unique to the *melanogaster* clade. Species of the *yakuba* clade (*D. yakuba*, *D. santomea*, and *D. teissieri*) exhibit a

small outgrowth in a similar position to the posterior lobe that may be homologous to the posterior lobe (Yassin & Orgogozo, 2013). Phylogeny is based upon that of Prud'homme (Prud'homme et al., 2006).

2.2.2 An Ancestral Enhancer of *Pox Neuro* was Co-opted into the Posterior Lobe Network

To trace the evolutionary history of the posterior lobe, we first examined *Pox neuro* (*Poxn*), a gene that is critical to its development. *Poxn* encodes a paired-domain transcription factor required for proper posterior lobe formation (Boll & Noll, 2002). In a comprehensive survey of the regulatory region of *Poxn*, a segment spanning the second exon and intron (Figure 2.3A) was found to be required for posterior lobe development (Boll & Noll, 2002). To examine the role of this enhancer in genital development and identify how this role evolved, we cloned this segment of the *D. melanogaster Poxn* gene into a GFP reporter construct (Figure 2.3A). Transgenic animals bearing the genital enhancer of *Poxn* drive expression both before and during posterior lobe development. At 32 hours after puparium formation (hAPF), a time that precedes the formation of the posterior lobe (see Figure 2.4A-L for a time course of genital development in lobed and non-lobed species), we observed broad GFP expression in a zone that straddles the presumptive clasper and lateral plate (Figure 2.3D). As the posterior lobe emerges from the lateral plate, and assumes its adult morphology, the reporter expresses high levels of GFP in the developing lobe (Figure 2.3E, arrow). This portion of the *Poxn* regulatory region accurately recapitulates the endogenous expression of *Poxn* mRNA and protein in the *D. melanogaster* lateral plate (Figure 2.3B-C; Figure 2.4M-N).

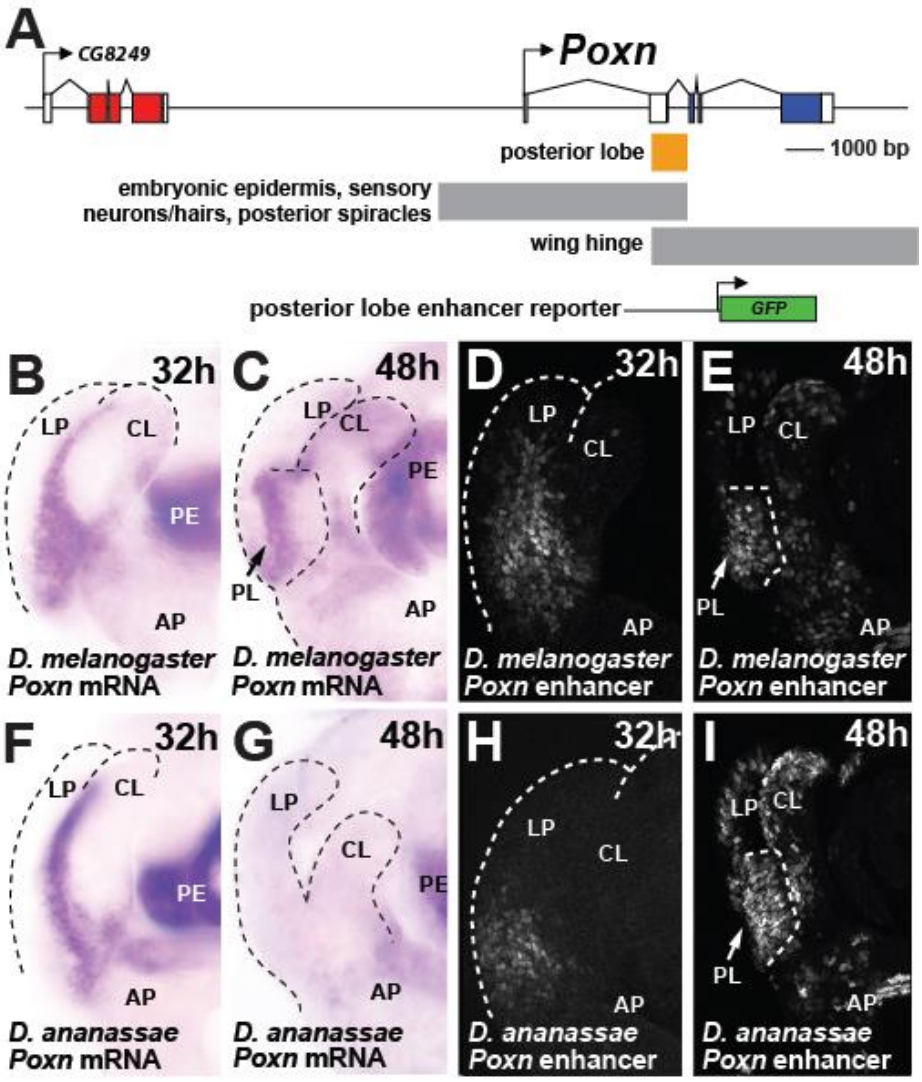

Figure 2.3. A Deeply Conserved Enhancer of *Poxn* is Required for Posterior Lobe Development.

(A) Schematic of the *Poxn* locus, displaying a subset of the described enhancer activities (Boll & Noll, 2002), and indicating the relative position of a posterior lobe reporter construct. (B, C) Accumulation of *Poxn* mRNA during genital development of *D. melanogaster* at (B) 32 hAPF (C) and 48 hAPF. (D, E) Activity of the *D. melanogaster* posterior lobe reporter at (D) 32 hAPF and (E) 48 hAPF. (F-G) Expression of *Poxn* in *D. ananassae* showing expression in the region between clasper and lateral plates (F), but not at the site where a lobe would develop (G). (H, I) Despite the absence of a posterior lobe in *D. ananassae*, the orthologous posterior lobe enhancer region drives expression preceding (H) and during posterior lobe development of *D. melanogaster* (I). CL clasper; LP lateral plate; AP anal plate, PE penis, PL posterior lobe.

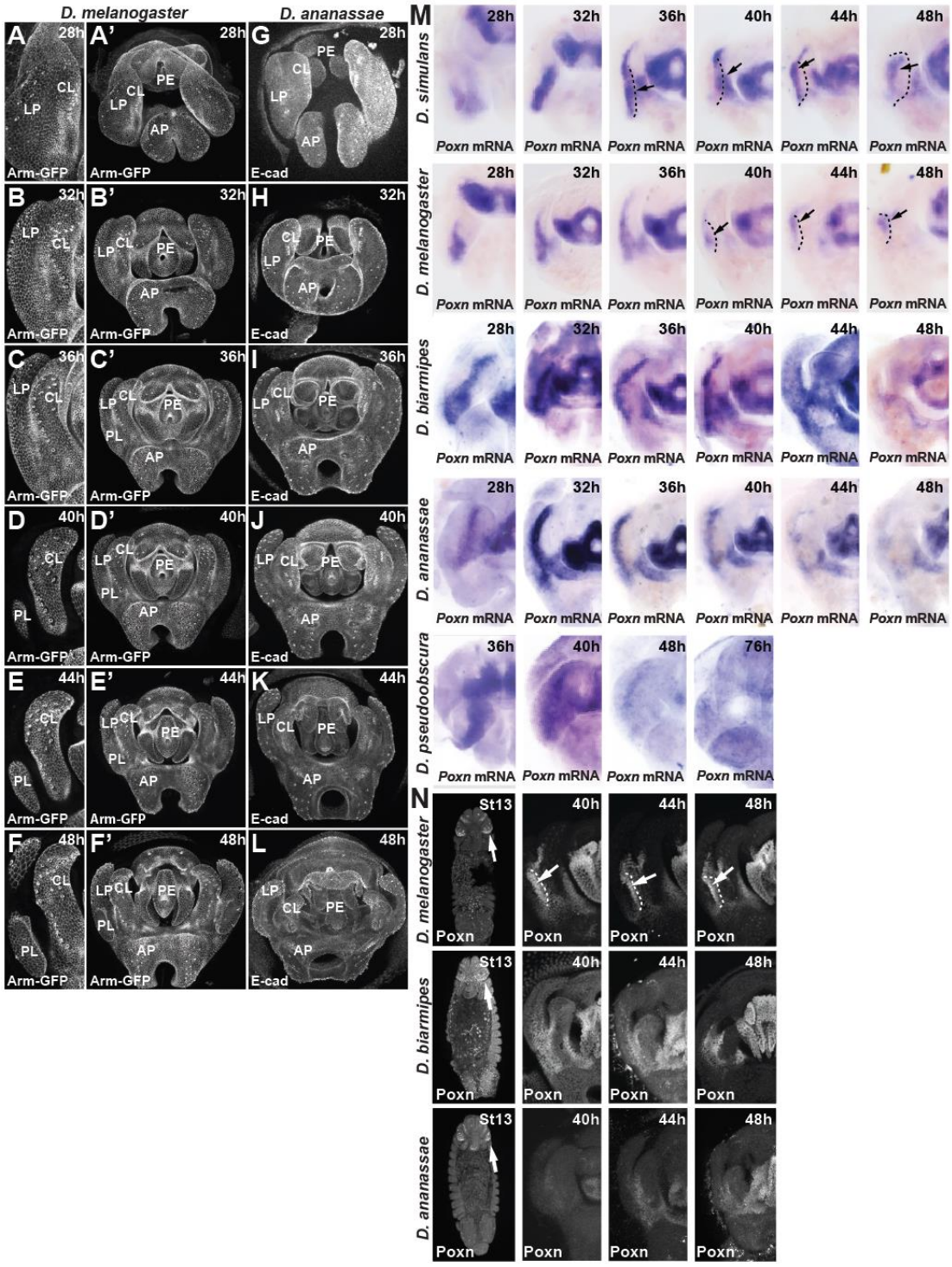

Figure 2.4. Comparison of Epithelial Morphogenesis and *Poxn* Expression Between Lobed and Non-lobed Species.

(A-F, A'-F') Confocal images of timed pupal genitalia of an *armadillo*-GFP line highlight changes in epithelial arrangement by monitoring apical cell junctions. (A) At 28 hAPF, the lateral plate (LP) and clasper (CL)

form a continuous epithelium. (B) At 32 hAPF, a furrow divides the continuous epithelium into presumptive lateral plate and clasper formations. (C, D) By 40 hAPF, the clasper and lateral plate are completely separated, resembling their adult structures, and the posterior lobe (PL) is visible as a ridge of cells emerging from the lateral plate epithelium. (E, F) by 44 hAPF, the posterior lobe has emerged from the lateral plate, and is in the process of adopting its adult shape. (A'-F') The anal plate (AP) and penis (PE) are adjacent to the lateral plate and clasper epithelial tissue. (G-L) Confocal images of timed *D. ananassae* pupal genitalia stained with antibody specific for E-cadherin from 28h to 48h demonstrate that these developmental processes are largely conserved except for the development of the posterior lobe. (M) *Poxn* mRNA is expressed during clasper and lateral plate cleavage in lobed and non-lobed species, but persists in lobed species at the base of the posterior lobe. *D. melanogaster* and *D. simulans* develop posterior lobes, *D. biarmipes, D. ananassae and D. pseudoobscura* do not. All species exhibit *Poxn* expression during clasper and lateral plate cleavage 28 hAPF to 40h APF, (36 hAPF to 48 hAPF for *D. pseudoobscura*). Only lobed species express *Poxn* during later genital development as the posterior lobe emerges 40 hAPF to 48 hAPF (black arrows, dashed lines). (N) (left) Poxn protein is expressed in a conserved pattern in the posterior spiracles of *D. melanogaster, D. biarmipes,* and *D. ananassae.* (right) Poxn protein is highly expressed at the base of the lobe of *D. melanogaster*, but only weak levels of Poxn protein persist, perduring from the earlier phase of expression in the non-lobed species *D. biarmipes* (middle) and *D. ananassae* (bottom).

The high level of reporter and *Poxn* mRNA in the developing posterior lobe strongly suggests that *Poxn* plays a direct role during posterior lobe development. To examine how this role evolved, we first analyzed its expression in species that lack this structure. At 32 hAPF, the early pattern of *Poxn* expression in the non-lobed species *D. ananassae* greatly resembles that of *D. melanogaster* prior to posterior lobe formation (Figure 2.3F). However, *Poxn* expression quickly subsides in the *D. ananassae* lateral plate once it has separated from the clasper (Figure 2.3G). Similar results were obtained for two additional non-lobed species, *D. biarmipes* and *D. pseudoobscura* (Figure 2.4M-N), suggesting that late, high levels of *Poxn* expression are uniquely associated with the development of this novelty.

Differences in *Poxn* expression between lobed and non-lobed species may be due to changes in the posterior lobe enhancer region (i.e. in *cis*), or could be caused by changes in *trans* that altered upstream regulators in the genitalia (Wittkopp, 2005). To distinguish between these possibilities, and ascertain whether the posterior lobe enhancer of *Poxn* recently derived its function, we examined the activity of this enhancer from species that lack this structure. Sequences orthologous to the *D. melanogaster* posterior lobe enhancer region were cloned from several non-lobed species, and tested for the ability to drive GFP reporter expression in the *D. melanogaster* posterior lobe. The posterior lobe enhancer regions of *D. ananassae*, *D. yakuba*, and *D. pseudoobscura Poxn* all drove GFP expression that closely matched the pattern and timing of the *D. melanogaster* reporter construct (Figure 2.3H-I; Figure 2.5G'-I'). The ability of the posterior lobe enhancer region to produce strong expression in the developing posterior lobe, despite the lack of this structure in these species strongly indicated that it predated the evolution of this novelty.

As our findings implied the absence of functionally significant changes in the *Poxn* enhancer during the evolution of the posterior lobe, we next tested whether a non-lobed species' enhancer could rescue the posterior lobe of a *D. melanogaster Poxn* mutant. The *D. melanogaster* posterior lobe enhancer is capable of generating a mild rescue of the *Poxn* null posterior lobe phenotype when fused to Gal4, driving a UAS-Poxn construct (Figure 2.5N). We observe that the orthologous regulatory region of *D. pseudoobscura* is also capable of generating a similar degree of rescue (Figure 2.5O). These experiments confirm the ancestral capability of the posterior lobe enhancer region to drive the expression necessary to generate a derived structure, suggesting that an ancestral function of this region was co-opted during the evolution of this novelty. We subsequently considered what this ancestral activity may be.

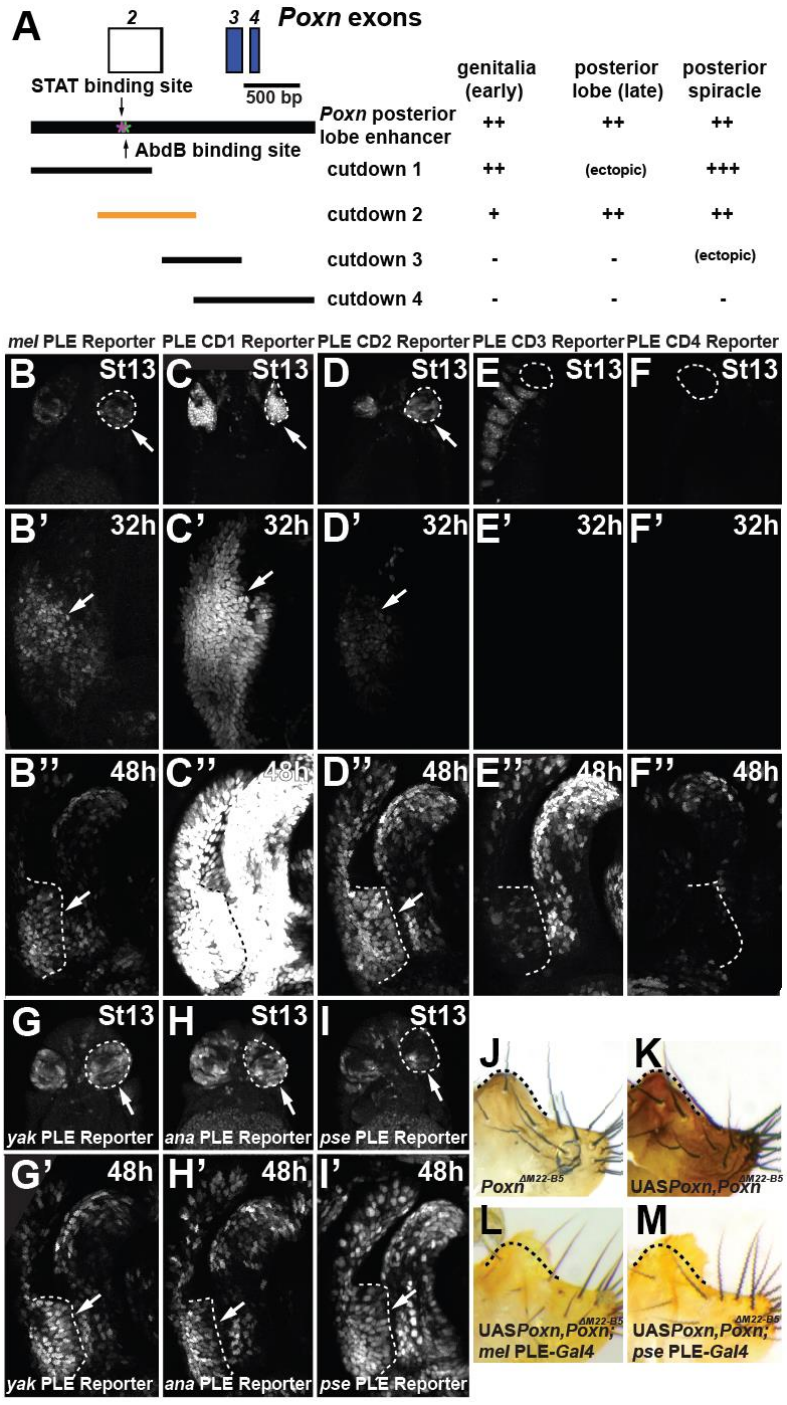

Figure 2.5. The *Poxn* Posterior Lobe Enhancer (PLE) is Inseparable from the Posterior Spiracle Enhancer and Pre-dates the Evolution of the Posterior Lobe.

(A) Subdivision of the posterior lobe enhancer region. All three activities encoded in this region can be isolated to the smaller cutdown 2 (CD2) truncation construct. Expression designation for each construct is listed

(present (+), ectopic, or absent (-)). (B-F) Expression of *Poxn* reporter constructs in stage 13 embryos, 32 hAPF genital samples (B'-F'), and at 48 hAPF (B"-F"). Compared to the full *Poxn* PLE reporter, the CD2 region provides the most faithful recapitulation of the larger segment (D-D"). (G-I, G'-I') Reporters of the orthologous regulatory regions of non-lobed species *D. yakuba, D. ananassae* and *D. pseudoobscura* exhibit expression during posterior spiracle development at stage 13 (G-I) and posterior lobe development at 48 hAPF (G'-I'). (J-M) *Poxn* mutant flies (J) fail to form a posterior lobe ($Poxn^{\Delta M22\text{-}B5}$). A UAS-*Poxn* construct (K) is not able to rescue the posterior lobe in the absence of a GAL4 driver (*UAS-Poxn* , $Poxn^{\Delta M22\text{-}B5}$), but when driven by the *Poxn* PLE fused to GAL4 (L), the posterior lobe is partially rescued (*UAS-Poxn*, $Poxn^{\Delta M22\text{-}B5}$*; mel PLE-GAL4*). (M) The *Poxn* PLE from the non-lobed species *D. pseudoobscura* is also able to partially rescue the posterior lobe of a *Poxn* mutant (*UAS-Poxn*, $Poxn^{\Delta M22\text{-}B5}$*; pse PLE-GAL4*). Dashed lines mark the contour of the $Poxn^{\Delta M22\text{-}B5}$ mutant lateral plate for reference.

In the initial screen of the *Poxn* regulatory region (Boll & Noll, 2002), several additional activities of *Poxn* were mapped to a domain overlapping the posterior lobe activity (Figure 2.3A). As these specificities may represent ancestral functions that were co-opted as the posterior lobe originated, we examined whether any of these were contained within our reporter fragment. Although many of the described activities were located outside of our reporter construct, strong expression was observed in an embryonic structure, the posterior spiracle (Figure 2.6A). Indeed, further subdivision of our reporter fragment failed to separate posterior spiracle from posterior lobe activities (Figure 2.5A). We next evaluated the possibility that the posterior spiracle enhancer of *Poxn* was co-opted during the origination of the posterior lobe.

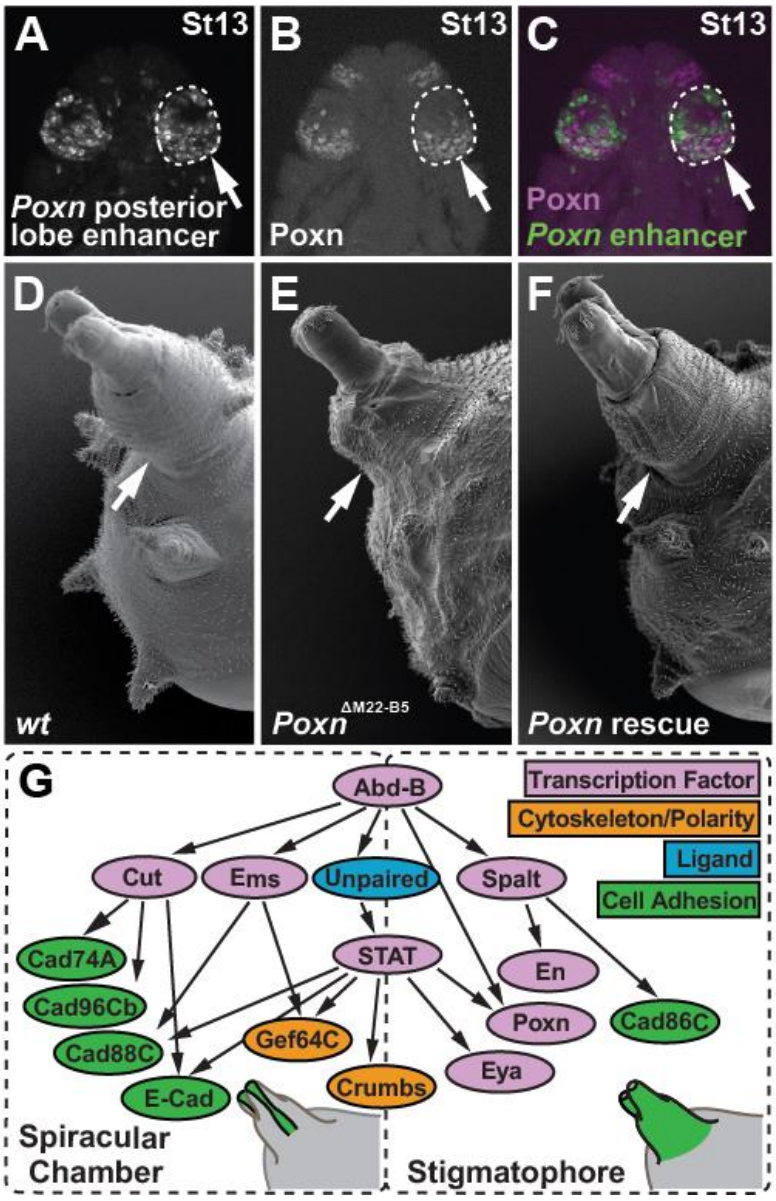

Figure 2.6. The Posterior Lobe Enhancer of *Poxn* is Active in the Hox-Regulated Network of the Posterior Spiracle.

(A) Transgenic embryo bearing the *D. melanogaster* posterior lobe enhancer reporter. (B) Antibody staining of Poxn protein in the posterior spiracle anlagen of the stage 13 (St13) *D. melanogaster* embryo presented in panel A. (C) Merged image of panels A and B, showing the *Poxn* enhancer (green) and Poxn protein (magenta). (D) Scanning electron micrograph of a wild-type third instar larva, showing the posterior spiracle structure. (E) A *Poxn* null mutant posterior spiracle is shorter relative to wildtype. (F) Rescue of posterior spiracle defects of a *Poxn* mutant by a fragment of the *Poxn* locus containing the lobe/spiracle enhancer fused to a P*oxn* cDNA. (G) Diagram

of posterior spiracle network, adapted from (Hu & Castelli Gair Hombría, 1999; Lovegrove et al., 2006). The addition and placement of *Poxn* and *eya* within this network is based upon data presented in this work. Arrows in A-F point to the posterior spiracle.

2.2.3 Shared Topology and Membership of the Posterior Lobe and Spiracle Networks

The posterior spiracle is a larval structure that is connected to the tracheal system, providing gas exchange to the larva (Figure 2.6D). *Poxn* is expressed in the embryonic region that develops into the posterior spiracle (Figure 2.6B), and *Poxn* mutants exhibit multiple defects in the spiracle, including transformation of sensory structures (Boll & Noll, 2002), and a shortening of the stigmatophore, an external protuberance that supports the spiracle (Figure 2.6E). The stigmatophore defect of *Poxn* can be rescued by a transgenic construct containing the posterior lobe and spiracle enhancer fused to a *Poxn* cDNA (Figure 2.6F). The posterior spiracle is specified during embryogenesis by a network of genes that is activated by the Hox gene *Abdominal-B (Abd-B)* (Figure 2.6G) (Hu & Castelli Gair Hombría, 1999). Intriguingly, genital development also depends upon *Abd-B*, resulting in genital-to-leg transformations in its absence (Estrada & Sánchez-Herrero, 2001).

Considering the apparent parallels between the posterior lobe and the posterior spiracle, we speculated that additional components of the spiracle network might be active in the developing genitalia. The JAK/STAT pathway plays a critical role in the posterior spiracle network (Lovegrove et al., 2006), and its ligand, encoded by the *unpaired* gene (*upd,* also known as *os*) (Harrison, McCoon, Binari, Gilman, & Perrimon, 1998), is expressed at high levels in the developing posterior lobe (Figure 2.7A). This pattern is consistent with the activity of a

JAK/STAT signaling reporter (Bach et al., 2007) , which is expressed at high levels during posterior lobe development (Figure 2.7B, Figure 2.8D-F). Reduction of JAK/STAT signaling in the genitalia by transgenic RNAi hairpins directed towards the receptor (*dome*), kinase (*hop)* or transcription factor (*Stat92E*) resulted in drastic reductions in the posterior lobe's size compared to a control RNAi hairpin (Figure 2.7C-F, 2.7T). Hence, the major signaling pathway that patterns the posterior spiracle is also active in the novel posterior lobe structure.

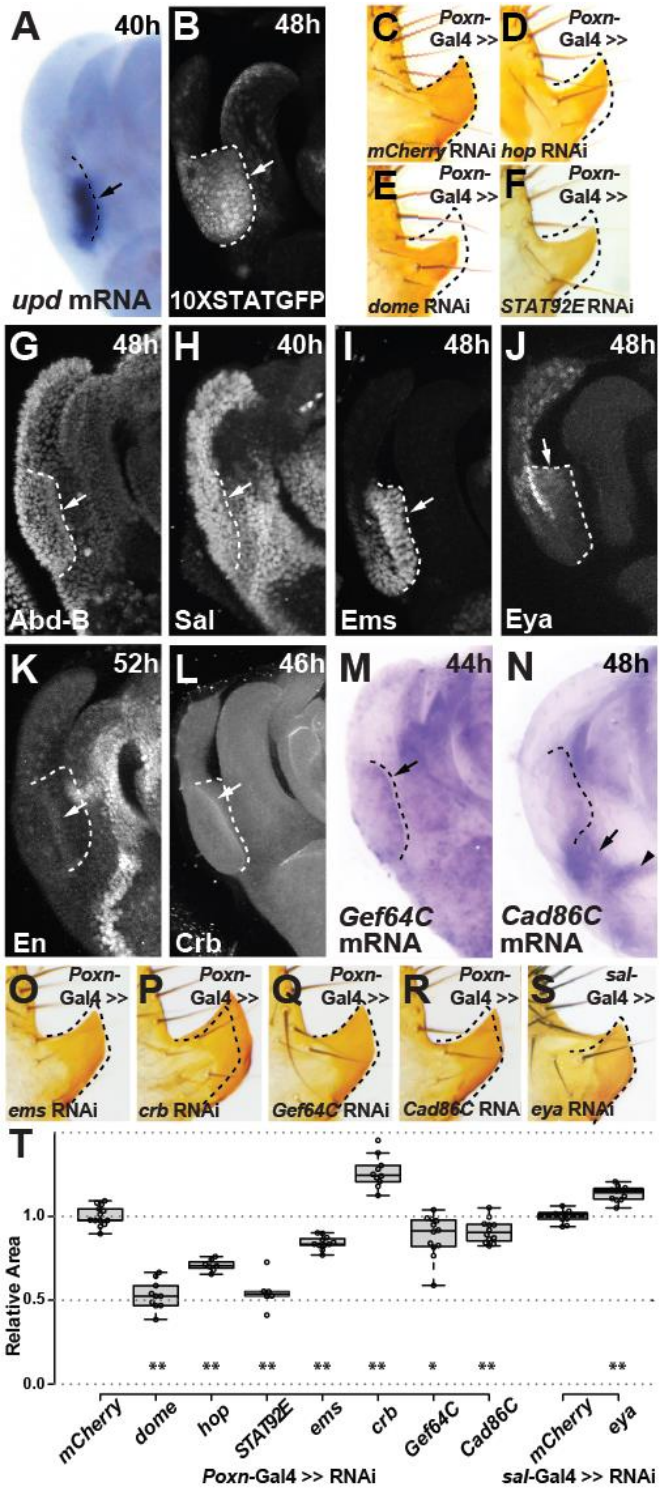

Figure 2.7. Shared Topology and Membership of the Posterior Lobe and Spiracle Networks.

Antibody staining (G-L) and *in situ* hybridization (A, M, N) reveal the deployment of several posterior spiracle network genes within the posterior lobe during genital development (arrows). (A-B) Expression of *upd*

mRNA in the developing lobe (A) closely mirrors the activity of a 10XStat92E-GFP reporter (B). (C-F) Reduction in expression of members of the JAK/STAT signaling pathway *hop* (D), *dome* (E) or *Stat92E* (F) reduces the size of posterior lobe compared to a control (C). (I-H) The top-tier spiracle network factor Ems (I) is strongly expressed within the developing posterior lobe, while Abd-B (G) and Sal (H) are present more generally throughout the lateral plate from which the lobe emerges. (J-N) Downstream spiracle network factors Eya (J) and *en* (K), as well as terminal differentiation factors Crb (L), *Gef64C* (M), and *Cad86C* (N) are all expressed at specific regions and stages of posterior lobe development (Figure 2.9). (O-S) Transgenic RNAi hairpin mediated reduction in expression of spiracle network members *ems* (D), *crb* (E), *Gef64C* (F), *Cad86C* (F) or *eya* (F) alters the size of posterior lobe compared to a control (shown in panel C). (T) Box plot depicting the relative area of posterior lobes upon RNAi treatments normalized to a control. Asterisks denote significance difference from control (student's paired t-test, *p <.05, ** p <.005). Dashed lines mark the position of the developing posterior lobe. (A, B, G-N) or demonstrate altered posterior lobe shape (D-F, O-S) compared to a control (C). Arrowhead in (N) identifies pattern not unique to lobed species (Figure 2.9E).

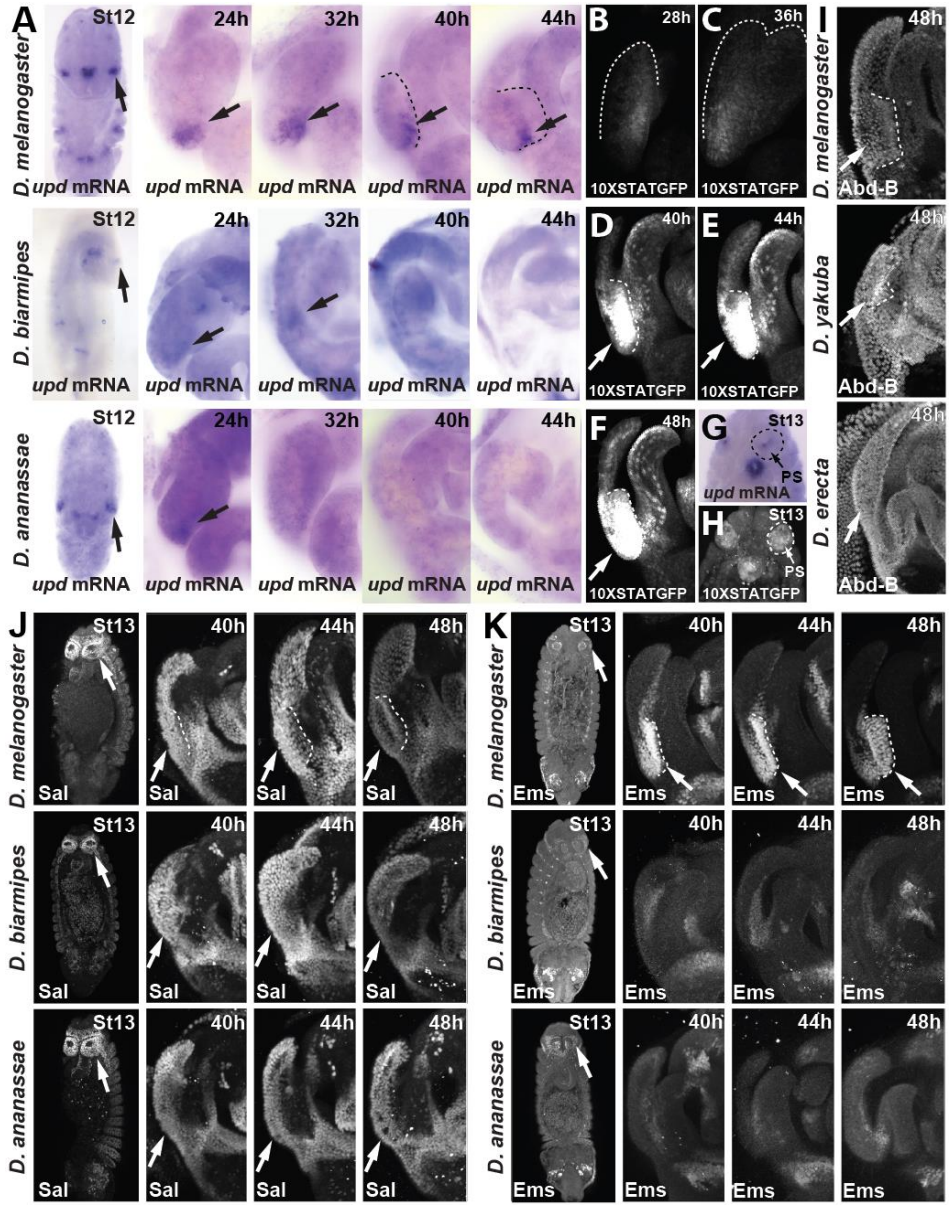

Figure 2.8. The Role of JAK/STAT Signaling and Top Tier Posterior Spiracle Network Factors Abdominal-B (Abd-B), Spalt (Sal) and Empty Spiracles (Ems) in Posterior Lobe Development.

(A) (left) mRNA of the JAK/STAT ligand *upd* accumulates in a conserved pattern in the embryonic posterior spiracle at stage 12 (right). In *D. melanogaster* (top), *upd* mRNA is steadily produced (arrows) both preceding (24h) and during the emergence of the posterior lobe (40-44 h, dashed lines). In the non-lobed species, *D. biarmipes* (middle), *upd* mRNA is faintly visible in a region reminiscent of the early phase of *upd* expression in *D. melanogaster*. In *D. ananassae* (bottom), the second non-lobed species, a similar early pattern of faint *upd* mRNA accumulation is visible. (B-F) A reporter containing 10 multimerized STAT92E binding sites (Bach et al., 2007)

reveals JAK/STAT signaling during cleavage of the clasper from the lateral plate 28 hAPF to 36 hAPF, and during posterior lobe development 40 hAPF to 48 hAPF (arrows), mirroring the expression of *upd* mRNA. (G) *in situ* hybridization visualizing mRNA for the JAK/STAT ligand *unpaired* in a stage 13 embryo. (H) 10XSTAT reporter reveals JAK/STAT signaling is active throughout the stigmatophore portion of the posterior spiracle (PS). (I) Except for the anal plate, Abd-B is expressed throughout the entire genitalia of both lobed (*D. melanogaster*, top) and non-lobed species (*D. yakuba*, middle; *D. erecta*, bottom) pupal genitalia at 48 hAPF. (J, K) Sal (J) and Ems (K) expression patterns. Stage 13 embryos (left) and timed pupal genitalia (right) are shown. (J) Spalt is expressed in a conserved pattern within the entire lateral plate of both lobed *(D. melanogaster*, top) and non-lobed species (*D. biarmipes*, middle; *D. ananassae*, bottom). (K) While Ems has a highly conserved pattern in the embryonic posterior spiracle, its genital expression differs greatly between lobed and non-lobed species. In *D. melanogaster* (top), Ems is highly expressed in the developing posterior lobe. In non-lobed species *D. biarmipes* (middle), and *D. ananassae* (bottom), Ems is only expressed at early stages in the region that will bisect the presumptive clasper and lateral plate, patterns that are visible at 40-44 hAPF.

We identified three additional top-level transcription factors of the posterior spiracle network that are active during the development of the posterior lobe. Abd-B and Spalt proteins are both deployed in broad domains that include the posterior lobe (Figure 2.7G-H; Figure 2.8D-E), consistent with severe genital defects in *Abd-B* (Estrada & Sánchez-Herrero, 2001; Foronda, Estrada, de Navas, & Sanchez-Herrero, 2006) and *spalt* mutants (Dong, Todi, Eberl, & Boekhoff-Falk, 2003). In contrast, Empty spiracles (Ems), named for its spiracle phenotype (Jürgens, Wieschaus, Nüsslein-Volhard, & Kluding, 1984), is expressed in a restricted genital pattern similar to *Poxn* (Figure 2.7I; Figure 2.8A). In summary, five transcription factors required for posterior spiracle development (Abd-B, Poxn, Spalt, Ems, and activated STAT) are deployed in the novel posterior lobe context, suggesting a highly similar *trans* regulatory landscape governing these two structures.

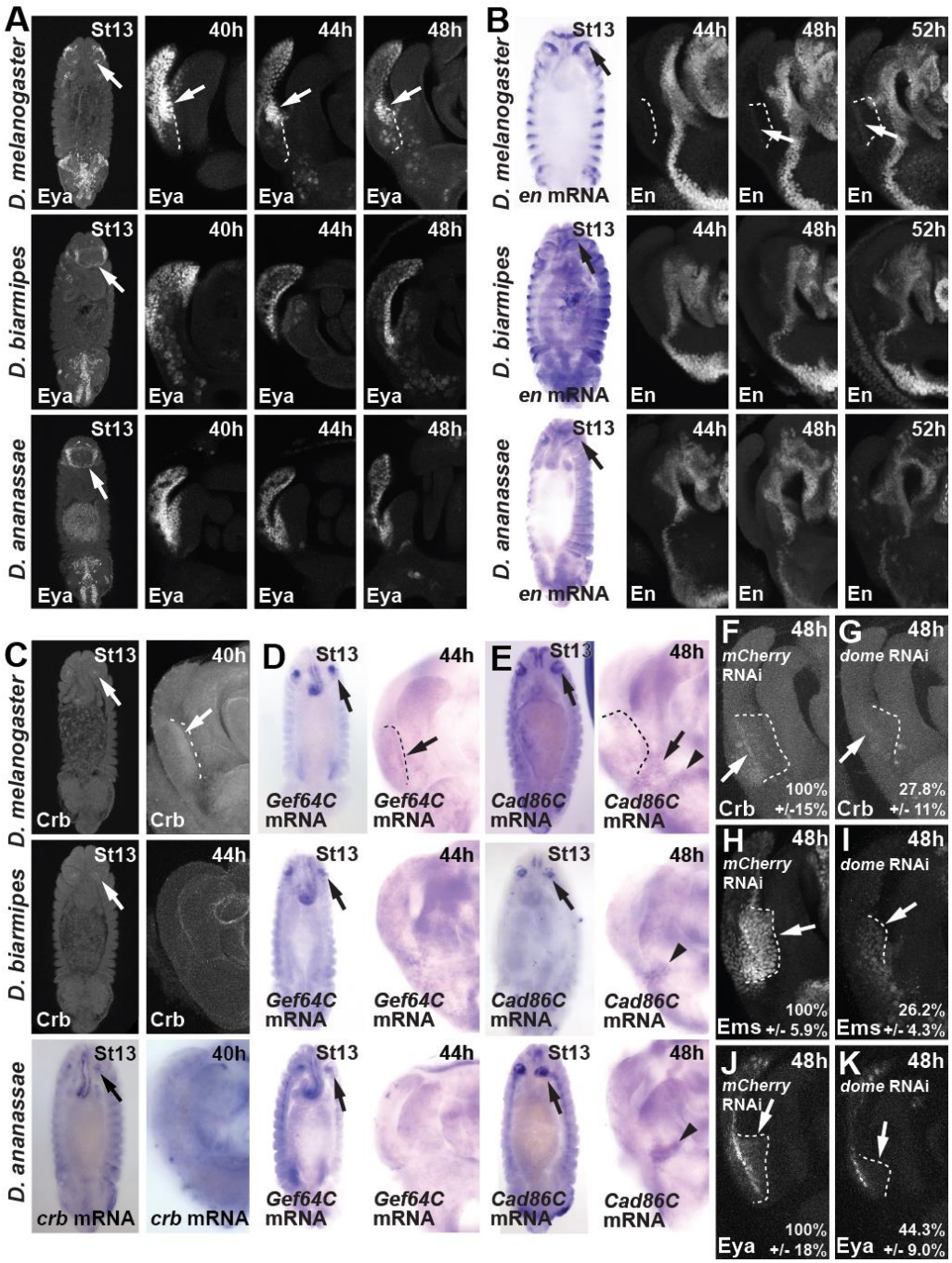

Figure 2.9. Comparative Expression Analysis of Downstream and Terminal Spiracle Network Genes Eya (A), En (B), Crumbs (C), *Gef64C* (D), and *Cad86C* (E).

(A) (left) Eya exhibits a conserved expression pattern in the outer border of the posterior spiracle of stage 13 embryos. (right) During genital development, Eya is expressed in the lateral plate of lobed and non-lobed species.

In the *D. melanogaster* posterior lobe, Eya is repressed in the dorsal portion of the lateral plate and strongly expressed in the ventral portion of the posterior lobe (arrows). (B) (left) *en* mRNA is expressed in the posterior portion of the posterior spiracle (arrows). (right) En protein is expressed in a highly conserved pattern marking the posterior compartment of the pupal genitalia. In *D. melanogaster*, En is weakly expressed within the posterior lobe (arrows), a pattern that is absent in non-lobed species. (C) Detection of Crumbs protein and *crb* mRNA in *D. melanogaster*, *D. biarmipes*, and *D. ananassae* shows a highly conserved pattern of expression in the posterior spiracle (left), but in the genitalia, expression at high levels in the posterior lobe region is specific to *D. melanogaster* (arrow). (D) *Gef64C* mRNA accumulates in a highly conserved pattern in the developing spiracle (left), but much like Crumbs, is specifically expressed in the posterior lobe of *D. melanogaster*. (E) mRNA pattern for *Cad86C* is conserved in the posterior spiracle (left) but lobe-associated expression is unique to the lobe-bearing species *D. melanogaster* (top, arrow). Arrowheads mark an anal plate associated pattern that appears in both lobed and non-lobed species. (F-K). RNAi hairpins targeted at JAK/STAT signaling component *dome* reduces the expression of Crb (G), Ems (I) and Eya (K) compared to controls (F, H, J). Values in panels F-K denote the staining intensity expressed as a percentage normalized to a control ± S.E.M.

While the *trans* regulatory landscapes of the lobe and spiracle bear an unexpected resemblance, they also appear to impart a high degree of spatial specificity. Abd-B is restricted to posterior body segments (Celniker, Keelan, & Lewis, 1989), while Poxn, Spalt, and Ems rarely overlap in expression (Dalton, Chadwick, & McGinnis, 1989; Dambly-Chaudiere et al., 1992; Kühnlein et al., 1994). The JAK/STAT pathway is recurrently deployed during development, but very few tissue settings would include all five factors. We therefore reasoned that downstream genes in the spiracle network might be activated in the developing posterior lobe. To test this possibility, we monitored their expression during genital development. In five genes of this network: *engrailed* (*en*), *crumbs* (*crb*), *Gef64C*, *Cad86C*, and *eyes absent* (*eya*), we found corresponding expression within the developing posterior lobe (Figure 2.7J-N; Figure

2.9A-E). Hence, a total of at least ten genes are shared between the two networks. We investigated the hierarchal relationship between several of the identified genes by targeting components of the JAK/STAT pathway using RNAi hairpins specific to *dome* (Figure 2.9M-T). Two genes significantly reduced in this background have been linked to JAK/STAT activity in the posterior spiracle, *crb* (Lovegrove et al., 2006) and *eya* (see below). Their reduction in response to the perturbation of a top-tier spiracle network factor supports a shared topology between the two networks.

The sharing of genes between the spiracle and lobe networks may be due to their recent recruitment to posterior lobe development, which would predict that their expression is specific to species that possess this structure. To determine whether the activity of these genes differs between lobed and non-lobed species, we examined their expression in non-lobed species at timepoints corresponding to stages in which the *D. melanogaster* lobe emerges. Ems exhibits strong lobe-specific activity that is absent in non-lobed species (Figure 2.8A), however both Spalt and Abd-B are widely and strongly expressed in all species tested (Figure 2.8D-E). *upd* mRNA is weakly present in early genitalia prior to lobe development in both lobed and non-lobed species, but persists and intensifies in *D. melanogaster* during lobe development (Figure 2.8A). Downstream spiracle network genes *eya, en, crb, Gef64C* and *Cad86C* are active in several locations within the genitalia, but all exhibit unique lobe-specific expression patterns (Figure 2.9A-E). Thus, of the ten shared genes that we have discovered, eight are unique to lobed species during the stages of this structure's emergence.

To confirm that the identified posterior spiracle genes actively participate in posterior lobe development, we targeted *ems, crb, Gef64C, Cad86C,* and *eya* with RNAi hairpins driven by genital drivers. Reduction of *ems, Gef64C and Cad86C* significantly reduced the size of the

posterior lobe, while reduction of *crb* and *eya* significantly increased the size of the lobe compared to a control RNAi hairpin (Figure 2.7O-T). Thus, genes of spiracle network that are specifically restricted to this novel structure during its development contribute to its construction.

2.2.4 Shared Enhancers Underlie the Parallel Topologies of the Lobe and Spiracle Networks

The striking similarity between the posterior lobe and spiracle networks may reflect the convergent evolution of similar network topologies, or could result from co-option of the ancestral posterior spiracle network in generating the lobe. To distinguish between co-option and coincidence, we tested additional enhancers of the posterior spiracle network for posterior lobe activity (see Experimental Procedures). In the case of co-option, multiple enhancers of the posterior spiracle network would be active in the posterior lobe, whereas convergence would produce enhancer activities in distinct locations within each shared gene's regulatory region.

The *crb* gene is deployed in the posterior spiracle through an intronic JAK/STAT responsive enhancer (Lovegrove et al., 2006), which we found to be active in the posterior lobe (Figure 2.10A, 2.10G, and 2.10G'). A recent screen of the regulatory regions of *invected (inv)* and *en* identified a posterior spiracle enhancer (Cheng et al., 2014), which consistently drives weak expression during late posterior lobe development (Figure 2.10B, 2.10H and 2.10H'). We discovered a region of the *Gef64C* gene that is active in both the posterior spiracle and posterior lobe patterns (Figure 2.10C, 2.10I, and 2.10I'). We also discovered a region of the *Cad86C* gene that consistently recapitulates a portion of its posterior spiracle expression domain, as well as a lobe-associated pattern that is specific to lobed species (Figure 2.10D, 2.10J and 2.10J', white arrow). For *eya*, a new member of the posterior spiracle network identified in this study (Figure

2.9A), we localized an upstream enhancer that recapitulates its genital expression pattern (Figure 2.10E and 2.10K'). This enhancer is also active in the outer edge of the larval spiracle's stigmatophore (Figure 2.10K). While a previously identified posterior spiracle enhancer upstream (US) of *ems* (Jones and McGinnis, 1993, Figure 2.10L) lacked activity in the posterior lobe (Figure 2.10L'), we identified an additional enhancer located just downstream of the transcription unit that is activated in both settings (Figure 2.10F, 2.10M, and 2.10M'). This downstream (DS) enhancer of *ems* recapitulates a previously undescribed activity in the outer edge of the stigmatophore (Figure 2.10N' and 2.10P') but is not active in the initial spiracular chamber pattern (Figure 2.10P). In conclusion, we have identified seven enhancers (*Poxn, crumbs, en, Gef64C, Cad86C, eya,* and *ems*) of the posterior lobe network that can be traced to overlapping functions in the posterior spiracle. Given the large size of their respective regulatory regions, we postulated that the coincidence of lobe and spiracle enhancers would be highly unlikely due to chance alone. Simulations in which we randomized the locations of lobe and spiracle reporter fragments across the full extent of each of the seven loci confirmed an extremely low probability that the observed lobe and spiracle enhancer fragments would overlap by a single nucleotide ($p = 6 \text{ X } 10^{-8}$).

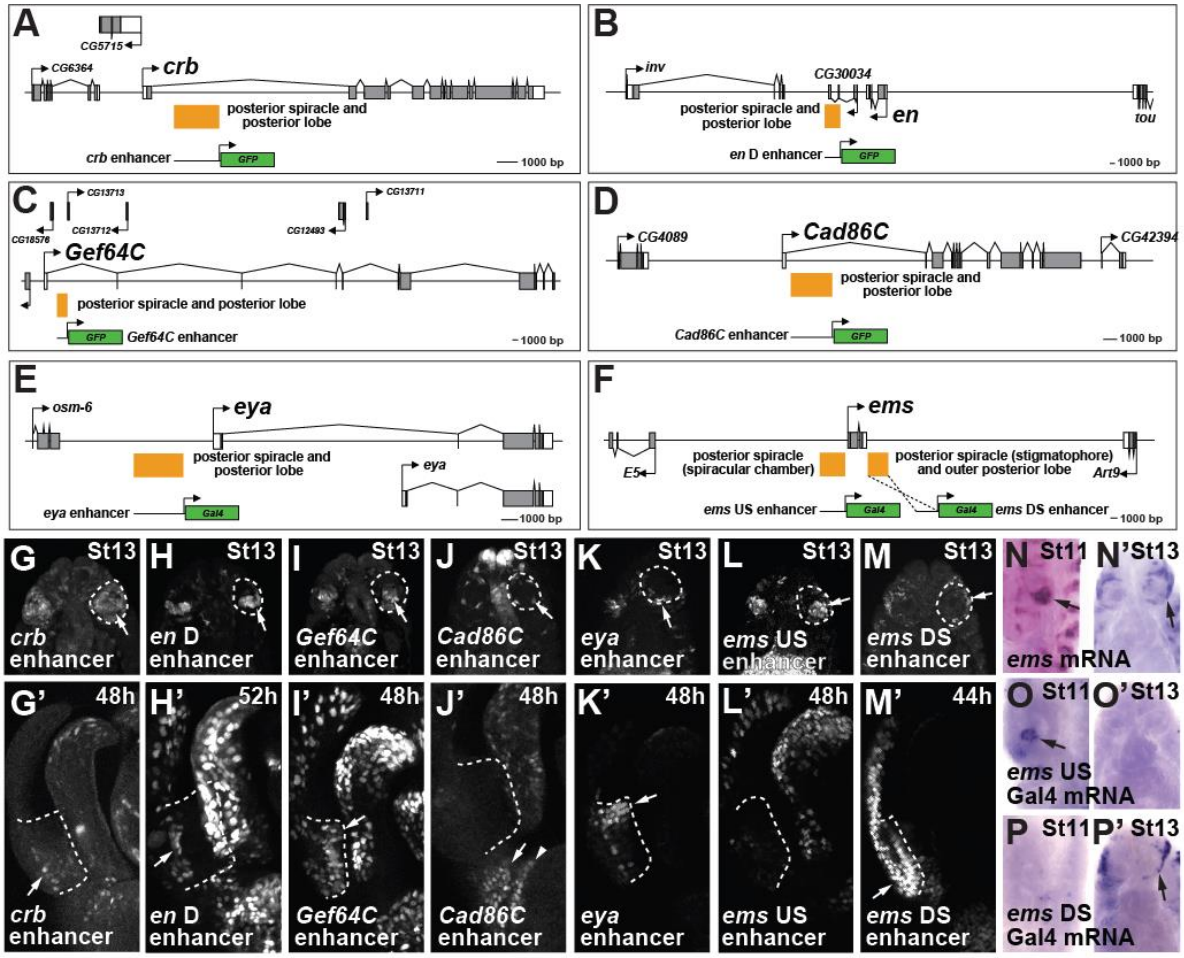

Figure 2.10. Co-option of Posterior Spiracle Enhancers to Posterior Lobe Development.

(A-F) Schematic diagrams of genomic loci in which an enhancer activated in both the posterior lobe and posterior spiracle were localized (orange boxes). Reporter constructs contain the schematized segment fused to either Green Fluorescence Protein (GFP) or Gal4. (G-M,G'-M') GFP reporter expression driven in transgenic *D. melanogaster* by enhancers for (G, G') *crb*, (H, H') *en*, (I, I') *Gef64C*, (J, J') *Cad86C*, (K, K') *eya*, (L, L') *ems* (US enhancer) and (M, M') *ems* (DS enhancer) in the posterior spiracle (G-M), and in the posterior lobe (G'-M'). (N-P) *ems* is first active at stage 11 in cells that contribute to the spiracular chamber, a pattern recapitulated by the *ems* US reporter (O), but not by the *ems* DS reporter (P). (N'-P') *ems* is also active later during posterior spiracle development around the border of the stigmatophore (arrow) and in each embryonic segment, a pattern not encoded in the upstream enhancer (O'), but is recapitulated by the *ems* DS reporter (P', arrow). (L') The *ems* US reporter is not expressed within the developing posterior lobe. (J') The *Cad86C* reporter also recapitulates a conserved pattern at the edge of the anal plate (arrowhead).

2.2.5 The Activation of Enhancers in Both New and Old Contexts Depends on Direct Input from Hox and Signaling Pathway Factors

A hallmark of co-option of regulatory sequences is the use of individual transcription factor binding sites in two or more developmental contexts (Rebeiz et al., 2011). The similarities in lobe and spiracle network topologies and enhancer locations strongly suggested that transcription factor binding sites within posterior spiracle enhancers would be required for posterior lobe function. Therefore, we searched for conserved transcription factor binding sites that could mediate functions common to both networks. Within the *Poxn* posterior lobe enhancer, we identified instances of high quality binding sites for STAT and Abd-B, both of which were contained within an 897 bp fragment active in both contexts (Figure 2.5A). In addition, we identified a high quality binding site for STAT within a 294 bp interval defined by two overlapping reporters of the *eya* enhancer that were active in both locations (Figure 2.11G and 2.11H). Comparisons to other sequenced *Drosophila* species revealed that these three sites are highly conserved (Figure 2.12A), consistent with their potential function in the deeply conserved posterior spiracle structure. Introduction of a 2-bp mutation that is known to disrupt STAT binding (Lovegrove et al., 2006) drastically reduced activity of the *Poxn* reporter in both the posterior lobe and posterior spiracle (Figure 2.12B-C and 2.12B'-C'), and similarly eliminated activity of the *eya* enhancer reporter in both settings (Figure 2.12E-F and 2.12E'-F'). Introduction of a 3-bp mutation that disrupts Abd-B binding (Williams et al., 2008) extinguished *Poxn* enhancer activity in the posterior lobe and significantly reduced posterior spiracle expression by 57% (Figure 2.12D and 2.12D'). Finally, we introduced mutations to three

conserved STAT binding sites known to disrupt activity of the *crb* enhancer in the posterior spiracle into our transgenic reporter system (Lovegrove et al., 2006), which eliminated expression of our *crb* reporter in the posterior lobe (Figure 2.12G-H). These results demonstrate the co-option of enhancers into a novel setting through the redeployment of pre-existing transcription binding sites.

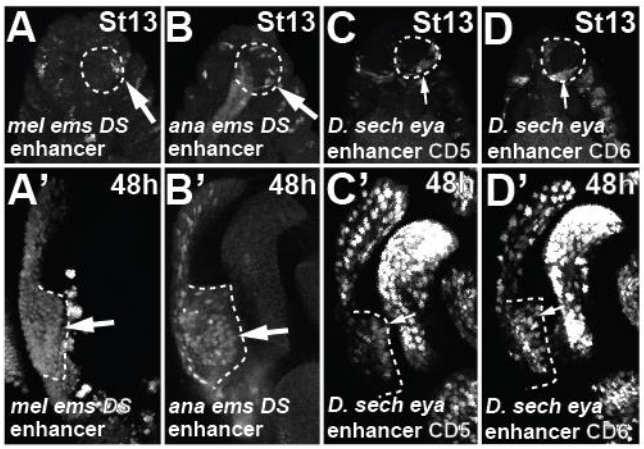

Figure 2.11. A Conserved Enhancer of *empty spiracles* (*ems*) was Co-opted to the Posterior Lobe, and the Spiracle and Lobe Activities of an *eyes absent* (*eya*) Enhancer are Inseparable.

(A-B, A'-B') Orthologous regions of the *ems DS* enhancer from lobed species *D. melanogaster* (A, A') and non-lobed species D. ananassae (B, B') both drive expression in the posterior spiracle (A, B) and posterior lobe (A', B'). (C-D, C'-D') Two overlapping subfragments of the *eya* spiracle/lobe enhancer cloned from *D. sechellia*, a lobed species, maintain both posterior lobe (C'-D') and posterior spiracle (C-D) activity (arrows).

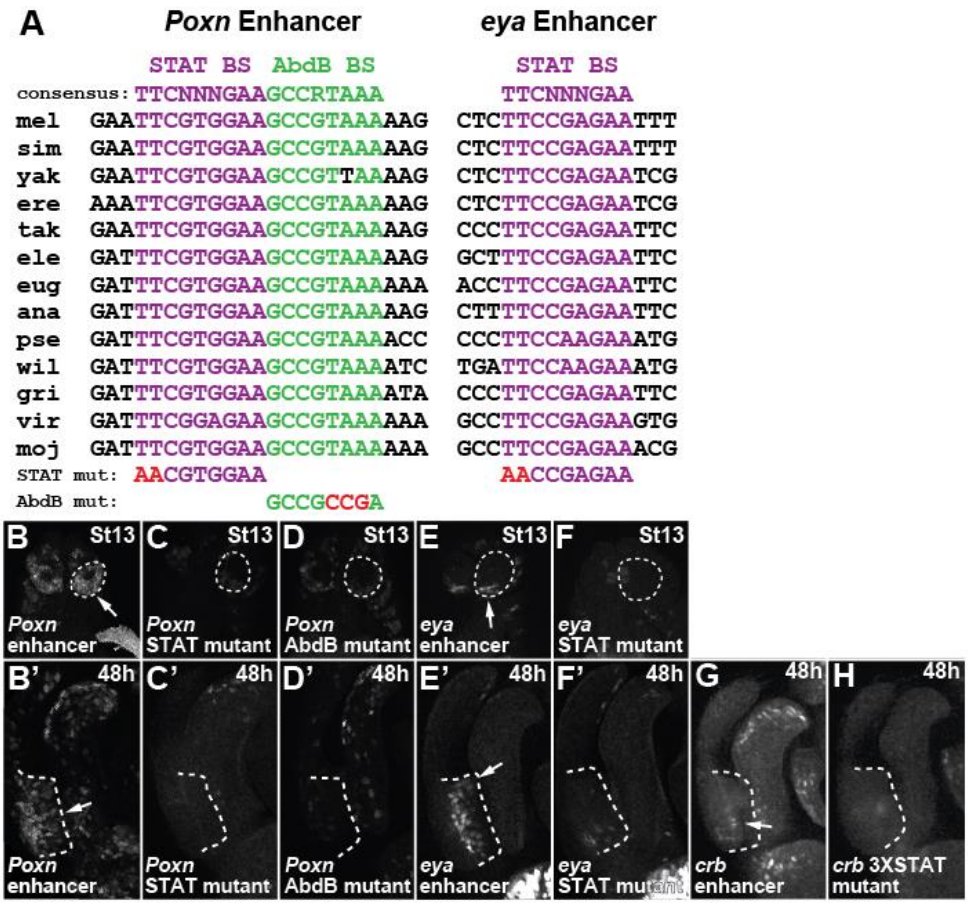

Figure 2.12. Redeployment of *crb, Poxn* and *eya* in the Posterior Lobe Required Ancestral Binding Sites for Abd-B and STAT that Function in the Posterior Spiracle Context.

(A) Alignment of a Stat92E binding site (purple text) and an Abd-B binding site (green text) of the *Poxn* lobe/spiracle enhancer and a Stat92E binding site (purple text) of the *eya* lobe/spiracle enhancer, showing near perfect conservation among sequenced *Drosophila* species. (B-D, B'-D') Mutations to two bases in a STAT binding site (C, C'), or three bases in an Abd-B binding site (D, D') reduces both posterior spiracle (C-D) and posterior lobe (C'-D') activity compared to the wildtype *Poxn* enhancer (B, B'). Mutation of two bases in a STAT binding site (F, F') reduces both posterior spiracle (F) and posterior lobe (F') activity compared to the wildtype *eya* enhancer (E, E'). (G-H) Mutations to three STAT binding sites known to disrupt activity of the *crb* lobe/spiracle enhancer in the posterior spiracle (Lovegrove et al., 2006), also eliminate activity in the posterior lobe (G) compared to the wildtype *crb* enhancer (H).

2.3 DISCUSSION

Here, we have shown how a gene regulatory network underlying a novel structure, the posterior lobe, is composed of components that are active in the embryonic posterior spiracle, an ancestral Hox-regulated structure that was present at the inception of this novelty (Fig. 2.13). These findings confirm previous speculation that network co-option proceeds through the re-use of individual transcription factor binding sites within enhancer sequences (Gao & Davidson, 2008; Monteiro & Podlaha, 2009). Further, our data help calibrate expectations concerning the degree of physical similarity between novel and ancestral structures during co-option events. Below, we briefly discuss how the architecture of the posterior spiracle network may have predisposed it for co-option in the genitalia, and explore the general implications of our findings with regard to the origins of morphological novelty.

While our results illustrate the downstream consequences of co-option, the upstream causative events await characterization. We suspect that some number of high-level regulators of the posterior spiracle network recently evolved novel genital expression patterns through alterations within their regulatory regions. Currently, Unpaired represents the best candidate upstream factor, as it is positioned near the top of the spiracle network, differs in expression greatly between lobed and non-lobed species (unlike Spalt and Abd-B), and is the only high-level factor in the spiracle network for which a shared lobe/spiracle enhancer has yet to be identified (Figure 2.13C). Indeed, a reporter screen of the 30kb of regulatory DNA immediately surrounding the *upd* gene identified a posterior spiracle enhancer that is not deployed in the posterior lobe, marking an important point of divergence separating the posterior spiracle and posterior lobe networks (Figure 2.13D-F).

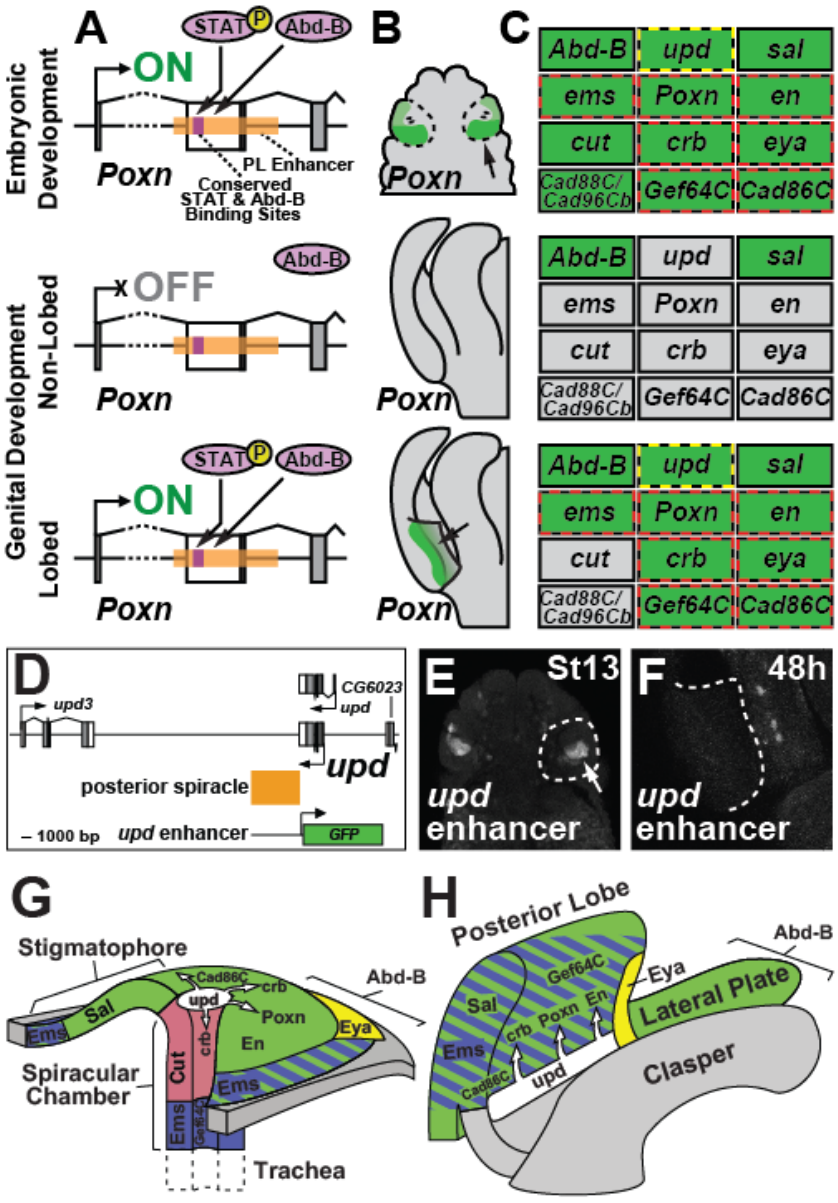

Figure 2.13. Model Depicting the Co-option of Genes, Enhancers, and Transcription Factor Binding Sites During the Origination of the Novel Posterior Lobe.

(A) (top) The posterior spiracle enhancer of *Poxn* binds Abd-B and phosphorylated STAT in the embryonic posterior spiracle anlagen to activate expression ("ON"). (middle) In species lacking a posterior lobe, the enhancer is not activated ("OFF"). (bottom) The deployment of regulatory factors of the spiracle network during late stages of genital development in lobed species resulted in the activation of the *Poxn* spiracle enhancer by Abd-B and activated

STAT. (B-C) Summary of *Poxn* expression (B) and the status of the posterior spiracle network (C) in the three developmental contexts. (C) Expressed genes are shaded in green, while inactive genes are shaded grey. Genes activated by a shared lobe/spiracle enhancer are outlined with red-dashes. The yellow dashes outlining the *upd* node indicates its activation in the spiracle through an enhancer that lacks lobe activity. (D) Schematic diagram of the *upd* locus in which a posterior spiracle enhancer was identified (orange box). (E-F) Reporter construct containing the schematized segment fused to Green Fluorescence Protein (GFP) is active in the posterior spiracle (E), but not in the posterior lobe (F).

The architecture of the posterior spiracle network may have shaped the possible developmental contexts in which it could be co-opted. The Hox factor Abd-B has a deeply conserved role in the insect abdomen and genitalia (Kelsh, Dawson, & Akam, 1993; Yoder & Carroll, 2006). The top-level factors of the posterior spiracle network depend upon *Abd-B* for activation in the embryo (Figure 2.6F) (Hu & Castelli Gair Hombría, 1999). This regulation by Abd-B extends to lower tiers of the network, such as *Poxn* (Figure 2.12A, 2.12D and 2.12D', Table 2.1). The tight integration of Abd-B with multiple tiers of the posterior spiracle network may have limited this network's re-deployment to posterior body segments that express Abd-B. Indeed, several components of this network (*Poxn, ems, upd*) are activated early during genital development in the presumptive cleavage furrow separating the lateral plate from the clasper (Figures 2.3B, 2.8A, K). This may represent the aftereffect of multiple waves of re-deployment in Abd-B expressing tissues. Examination of additional examples of network co-option at the level of constituent regulatory sequences could reveal general rules that govern and bias network redeployment.

Historically, the identification of co-option events has relied upon comparative analyses of gene co-expression. The first examples of co-option were diagnosed by finding novel gene

expression patterns near zones of ancestral function, such as the deployment of the posterior wing patterning circuit within novel butterfly eyespots (Keys et al., 1999). Subsequently, many examples of co-option have involved educated guesses of the types of networks that contribute to the novelty, such as the role of the appendage specification network within beetle horns (A P Moczek et al., 2006; A P Moczek & Nagy, 2005), or the sharing of the biomineralization network between adult and larval skeletons of sea urchins (Gao & Davidson, 2008). Our data suggest that tracing the evolutionary origins of individual enhancers provides a less biased path for connecting novelties to their ancestral beginnings, as any of the seven enhancers we have characterized in the posterior lobe would have led us to the spiracle network. Further, this approach is likely to illuminate the underlying cellular mechanisms by which the co-option of a network is translated into a novel developmental outcome.

Rather than generating a serial homolog of the posterior spiracle, the co-option event forming the posterior lobe resulted in an epithelial outgrowth, likely owing to the deployment of only a portion of the spiracle network in the genitalia. This is reflected by the absence of the Cut transcription factor and downstream genes (Figure 2.14B-L) that control the spiracular chamber's development (Hu & Castelli Gair Hombría, 1999). Of the ten genes we have identified in both networks, nine are active in the stigmatophore (Figure 2.13U-V), the outer sheath of the posterior spiracle that protrudes from the body through a process that involves convergent extension (Brown & Castelli Gair Hombría, 2000; Hu & Castelli Gair Hombría, 1999). Collectively, these findings imply that similar morphogenic processes are activated by this shared network in the novel setting of the posterior lobe. We propose that the inspection of enhancers underlying other novel three-dimensional structures may reveal similar networks that

have been used over and over again to generate “unique styles of architecture” within developing tissues (Zuckerkandl, 1976).

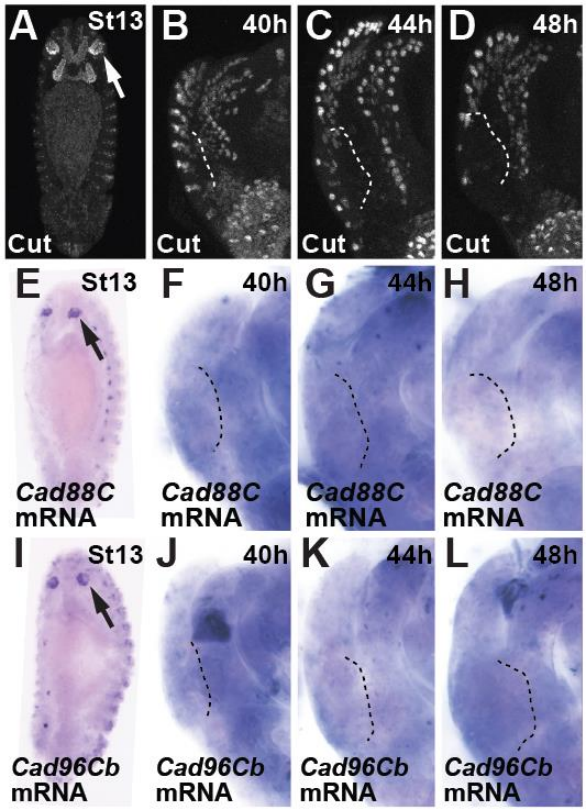

Figure 2.14. Spiracular Chamber Genes are not Expressed During Posterior Lobe Development.

(A-D) Cut antibody staining reveals expression in the spiracle (A) as previously reported (Hu & Castelli Gair Hombría, 1999), but in the developing genitalia, expression is limited to developing sensory organs (B-D). (E-H) The spiracular chamber gene *Cad88C* is expressed at high levels in the spiracle (e), but is absent during genital development (F-H). (I-L) A similar absence of *Cad96Cb* during posterior lobe development is observed. Dashed lines mark the position of the posterior lobe.

2.4 METHODS

2.4.1 Fly Strains and Husbandry

All flies were reared on a standard cornmeal medium. Species used in this study were obtained from the UC San Diego *Drosophila* Stock Center (*Drosophila biarmipes* #0000-1028.01, *Drosophila ananassae* #0000-1005.01, *Drosophila simulans* #14021-0251.165, *Drosophila pseudoobscura* #0000-1006.01, *Drosophila sechellia* #14021-0248.03, *Drosophila erecta* #14021-0224.01). The *Drosophila melanogaster* line used in this study is mutant for *yellow* and *white* (y^1w^1, Bloomington Stock Center #1495), and was isogenized for 8 generations.

2.4.2 Pupal Genital Sample Preparation

To collect developmentally staged genital samples, white prepupae were sorted by sex, and incubated at 25°C for 24 hours to 48 hours. Pupae were cut in half in cold PBS, extricated from the pupal case, and flushed with cold PBS to remove fat bodies and internal organs while preserving the developing genital epithelium. Carcasses were then fixed in PBS with 0.1% Triton-X and 4% paraformadehyde (PBT-fix) at room temperature for 30 minutes. Samples containing fluorescent reporters were washed three times for 10 minutes in PBS with 0.1% Triton-X (PBT) then imaged immediately. Samples to be used for *in situ* hybridization were rinsed twice in methanol and stored in ethanol at -20°C.

2.4.3 Embryo Collection

Embryos were collected from grape agar plates (Genesee Scientific) in egg-lay chambers that were incubated at 25°C for up to 20 hours. Embryos were dechorionated in 50% bleach for 3 minutes, washed in distilled water, and collected on a nitrile filter. Embryos were then fixed for 20 minutes in scintillation vials containing PBS, 2% paraformaldehyde, and 50% heptane. The PBS layer was removed from the vial and replaced with an equal amount of methanol. Samples to be used for *in situ* hybridization were vortexed for 30 seconds, removed from the methanol layer, rinsed twice in methanol then stored in ethanol. Samples containing fluorescent reporters or to be used for immunostaining were shaken vigorously by hand for 1 minute, rinsed in methanol once then quickly rinsed in PBT three times to prevent the degradation of GFP and antibody epitopes.

2.4.4 Immunostaining

Embryo and genital samples were incubated overnight at 4°C with primary antibodies diluted in PBT. The following primary antibodies were used: rabbit-anti-Poxn 1:100 (Dambly-Chaudiere et al., 1992), rabbit anti-Ems 1:200 (Dalton et al., 1989), rabbit anti-Spalt 1:500 (Barrio et al., 1996), mouse anti-Eya 1:100 (Bonini, Bui, Gray-Board, & Warrick, 1997), mouse anti-Crb 1:50 (Tepass & Knust, 1993), mouse anti-Engrailed/Invected 1:500 (Patel et al., 1989), rat anti-E-cadherin 1:100 (antibody DCAD2, Developmental Studies Hybridoma Bank), and mouse anti-Cut 1:100 (antibody 2B10, Developmental Studies Hybridoma Bank). After several washes with PBT to remove unbound primary antibody, samples were incubated overnight in diluted secondary antibody (donkey anti-mouse Alexa 488, and donkey anti-rabbit Alexa 647,

both at 1:400 dilution from Molecular Probes, or goat anti-rat Alexa 488 at 1:200 dilution from Molecular Probes) to detect bound primary antibody. Samples were washed in PBT to remove unbound secondary antibody, incubated for 10 minutes in 50% PBT and 50% glycerol solution, then mounted on glass slides in an 80% glycerol 0.1M Tris-HCL 8.0 solution.

2.4.5 *in situ* Hybridization

in situ hybridization was performed as previously described in (Rebeiz, Pool, Kassner, Aquadro, & Carroll, 2009) with the modification that we used an InsituPro VSi robot (Intavis Bioanalytical Instruments). Fixed embryo and genital samples were first dehydrated in a 50% xylenes/50% ethanol solution for 30 minutes at room temperature. Xylenes were removed by several washes with ethanol before the samples were loaded into the InsituPro VSi. During the automated steps, the samples were washed in methanol, rehydrated in PBT, fixed in PBT-fix, incubated in 1:25,000 proteinase K PBT (from a 10mg/mL stock solution), fixed in PBT-fix, and subjected to several washes in hybridization buffer. Samples were probed with a digoxygenin riboprobe targeting the coding regions of selected genes (primers listed in Table 2.1) for 18 hours at 65°C. Unbound riboprobe was removed in several subsequent hybridization buffer washes, and washed several times in PBT. Samples were removed from the robot, and incubated overnight in PBT with 1:6000 anti-digoxygenin antibody Fab fragments conjugated to alkaline phosphatase (Roche Diagnostics). Alkaline phosphatase staining was then developed for several hours in NBT/BCIP color development substrate (Promega). Samples were then washed in PBT and mounted on glass slides in an 80% glycerol 0.1M Tris-HCL 8.0 solution.

2.4.6 Transgenic Constructs

Enhancer elements were cloned using the primers listed in Table 2.2, and inserted into the vector pS3aG (GFP reporter) or pS3aG4 (Gal4 reporter) using *AscI* and *SbfI* restriction sites as previously described (Williams et al., 2008). Primers were designed and sequence conservation was assessed using the GenePalette software tool (Rebeiz & Posakony, 2004). Targeted regions were cloned from genomic DNA purified using the DNeasy Blood & Tissue Kit (Qiagen). Transcription factor binding site mutations were introduced using overlap extension PCR with mutant primers (Table 2.3). All GFP reporters were inserted into the 51D landing site on the 2nd chromosome (Bischof, Maeda, Hediger, Karch, & Basler, 2007), or the third chromosome 68A4 "attP2" site (Groth, Fish, Nusse, & Calos, 2004) by Rainbow Transgenics. Gal4 insertions depicted in Figure 2.5 were inserted into the 68E1 landing site on the third chromosome (Bischof et al., 2007). A full list of transgenes and insertions sites is listed in Table 2.4.

The *Poxn* rescue construct depicted in Figure 2.3E of the main text contains a 7.8kb genomic fragment containing 3kb upstream of the Poxn coding unit, including the *Poxn* promoter, and the first 3 exons and 2 introns of *Poxn* (which includes the lobe/spiracle enhancer). The remainder of the *Poxn* gene was joined to this construct from a *Poxn* CDNA. This construct ("L2") is identical to the "L1" construct published by Boll and Noll, but differs by the inclusion of 1.5 kb additional sequence upstream of the promoter (Boll & Noll, 2002).

The following GFP and Gal4 reporters were obtained from existing sources. 10XStat92E-GFP reporter was obtained from Erika Bach (Bach et al., 2007). Poxn-Gal4 (construct #13 from (Boll & Noll, 2002)) and UAS-Poxn was obtained from Werner Boll. *armadillo*-GFP was obtained from the Bloomington *Drosophila* Stock Center (#8556). Several enhancer-GAL4 lines from the Rubin collection (Pfeiffer et al., 2008) were obtained from the Bloomington *Drosophila*

Stock Center (BDSC) and are listed in Supplementary Table 2.4. Transgenic RNAi lines from the Harvard TRiP project include: *dome* (#34618), *Stat92E* (#33637) *hop* (#32966), *crb* (#40869), *Cad86C* (#27295), *Gef64C* (#31130), *ems* (#50673), *eya* (#35725). *mCherry* (#35785), a gene that is not present in the *Drosophila* genome was used as a control for RNAi experiments. The *salm*-Gal4 driver (#25755) was also obtained from the BDSC.

2.4.7 Microscopy

Adult posterior lobe cuticles and stained *in situ* hybridization samples were imaged on a Leica M205 stereomicroscope with a 1.6X objective with the extended multi-focus function. Samples stained with fluorescent antibodies or containing fluorescent reporters were imaged via confocal microscopy at 20X magnification on an Olympus Fluoview 1000 microscope. SEM images of third instar larvae were obtained as previously described by Higashijima (Higashijima, Michiue, Emori, & Saigo, 1992).

For each transgenic construct, 3-5 independent lines inserted into the 51D landing site (Bischof et al., 2007) or 68A4 “attP2” landing site (Groth et al., 2004) were derived. A list of reporters and corresponding landing sites are reported in Table 2.4. We compared the relative expression of multiple lines in the genitalia to determine the normal reporter activity of each construct. For quantitative measures, relative fluorescence of the *Poxn* and *eya* posterior lobe enhancers, and constructs mutant for STAT and Abd-B sites were determined in both the posterior lobe and posterior spiracle contexts. Mounted genital and embryo samples were imaged at 40X magnification under identical, non-saturating settings uniquely optimized for each sample

type. Relative expression within the lobe or spiracle was quantified using ImageJ and assessed using a student's paired t-test.

2.4.8 Simulations of Posterior Lobe and Spiracle Enhancer Co-occurrence

The lengths of shared enhancers and the length of each regulatory region in which these enhancers were embedded were input into an in-house Perl script, `CRE-overlap-sim`. This program randomizes the location of two equally sized segments of DNA (the size of each reporter fragment tested) across the length of each gene's potential regulatory sequence (the distance from the upstream gene to the gene downstream). For each simulation, the script measures whether the two segments overlapped, and counts a successful co-occurrence when all of the input enhancers overlap by the designated number of nucleotides in their respective regulatory regions. A large overlap, which would be expected for co-opted enhancer sequences will reduce the measured probability of co-occurrence. Our simulations specified a 1 nucleotide overlap, which represents the most permissive, and thus most stringent setting possible to detect non-random co-occurrence. 500,000,000 simulations were performed, and the average *p*-value as presented in the main text was calculated.

2.4.9 Identification of Shared and Distinct Posterior Spiracle/Posterior Lobe Enhancers

A combination of comprehensive whole gene surveys and targeted candidate region tests of non-coding regions of genes shared between the two networks was employed to identify co-opted enhancers. In the case of five out of eight of the identified enhancers, multiple constructs, inserted in at least two distinct genomic locations were tested for activity. For the whole gene

surveys, with the exception of *upd*, we used lines from the Rubin GAL4 collection (Jenett et al., 2012; Pfeiffer et al., 2008), in which non-coding sequences are fused to the GAL4 transcription factor, and inserted into the attP2 site on the third chromosome. We supplemented these searches by constructs we generated (see *"Transgenic Constructs"* above) when necessary. We detail the search for each of these enhancers below:

crb: Lovegrove et al. identified a spiracle enhancer located in the first intron (Lovegrove et al., 2006), for which we cloned an identical segment into our reporter system (Table 2.2). Additionally, we screened all intronic sequences using the Rubin-Gal4 collection, in which the construct overlapping the Lovegrove fragment uniquely recapitulated lobe expression. We additionally cloned the upstream region of *crb* into our GFP reporter system, and this fragment was not active in the posterior lobe.

en: Cheng *et al* screened the regulatory regions surrounding *engrailed* and *invected* (Cheng et al., 2014). The "D" enhancer from the intergenic region between *inv* and *en* was shown to specifically recapitulate the posterior spiracle activity of *en*. We reconstituted this enhancer by designing primers to clone the identical segment into our reporter system (Table 2.2). This construct drove strong expression similar to endogenous posterior spiracle *en* activity as reported by Cheng, and weak but consistent activity within a subset of the posterior lobe, mirroring the levels that appear late during posterior lobe development (Figure 2.9B).

eya: We screened the upstream region and introns of *eya* using the Rubin Gal4 collection. This screen identified a single posterior lobe activity just upstream of the transcription unit. We

further confirmed the activity of this enhancer fragment by inserting it into our GFP reporter system, which showed activity in the posterior lobe as well. Both the GFP reporter and the Rubin Gal4 constructs were expressed in the posterior spiracle. We further refined the size of this regulatory region by testing overlapping fragments of the *D. sechellia eya* enhancer. Two fragments that overlap by 294 bp were active in both spiracle and lobe tissues (Figure 2.11F-G'). The smallest fragment tested was 1060 bp.

ems: Rubin Gal4 lines existed for nearly all of the ~67 kb region encompassing the non-coding DNA surrounding *ems*. To test a portion of the regulatory region upstream of the *ems* promoter that is not included in the Rubin Gal4 collection, we cloned three additional overlapping regions into our GFP reporter system (Table 2.2). We first tested a Rubin Gal4 line that contains the previously identified upstream enhancer for the spiracular chamber (Jones & McGinnis, 1993) (Figure 2.10F). This line faithfully reproduced spiracular chamber expression (Figure 2.10L and 2.10O), but was not active in the *ems* posterior lobe pattern (Figure 2.10L'). Screening the other Rubin collection lines of *ems* for genital activity, we identified a fragment just downstream of the transcription unit that drove expression partially recapitulating the lobe expression of *ems*. To determine if this enhancer was indeed distinct from the posterior spiracle activity, we examined its expression in stage 13 embryos, and noticed that it was active in the outer stigmatophore (Figure 2.10M and 2.10P'), a pattern that recapitulates endogenous *ems* expression (Figure 2.10N'). We cloned a subfragment of this downstream enhancer into our GFP reporter system, confirming the activity of this segment in the posterior lobe and spiracle. In addition, we cloned the orthologous segment of DNA from *D. ananassae* into our reporter

system, demonstrating that a non-lobed species version of *ems* DS is capable of driving expression within the posterior lobe (Figure 2.11B').

Gef64C: A survey of the non-coding region of *Gef64C* identified a segment containing several binding sites for genes in the spiracle network, including a high affinity binding site for Abd-B (Ekker et al., 1994), and two candidate STAT binding sites, all of which were conserved to *D. pseudoobscura* (Table 2.5). Fusing this segment of DNA into our reporter system revealed expression in the spiracular chamber of the posterior spiracle, embryonic hindgut, and in several zones in the developing genitalia that recapitulate its endogenous expression (clasper, lobe, anal plate, and hypandrium, Figure 2.10I and 2.10I') and embryo. Further truncation of this segment of DNA separated the posterior spiracle and posterior lobe patterns from the other activities, localizing this enhancer to the first intron. This truncation includes the two candidate STAT binding sites but not the candidate Abd-B binding site (Table 2.5).

Cad86C: A screen of the non-coding regions surrounding *Cad86C* identified an intronic region near the promoter that included a Spalt site (Barrio et al., 1996) which is conserved to *D. ananassae* (Table 2.5). We cloned a 3003 bp segment of DNA that included this region into our reporter system. This reporter consistently recapitulated a portion of the endogenous *Cad86C* activity in the posterior spiracle and embryonic anus (Figure 2.10J and Figure 2.9E), and drove expression in the anal plate pattern common to both lobed and non-lobed species (Figure 2.10J' arrowhead and Figure 2.9E), as well as the lobe-specific pattern just posterior to the lobe (Figure 2.10J' arrow and Figure 2.9E).

upd: We screened the 30kb intergenic non-coding DNA between *upd* (also *os*) and its neighboring genes *upd3* and *CG6023* by cloning eight overlapping segments into our reporter system (Table 2.2). One reporter directly downstream of *upd* drove expression within the posterior spiracle (Figure 2.13E), matching the endogenous *upd* pattern (Figure 2.8G), and none of the tested reporters drove expression within the posterior lobe. The region that drove posterior spiracle expression contains a high quality match to the Abd-B binding site (Ekker et al., 1994), which is conserved to at least *D. virilis*.

2.4.10 Identification of Predicted Conserved Transcription Factor Binding Sites in Minimal Shared Enhancers

Using the GenePalette Software tool (Rebeiz & Posakony, 2004), we compared the orthologous regions of the shared posterior spiracle and posterior lobe enhancers from *D. melanogaster, D. simulans, D. yakuba, D. biarmipes, D. ananassae, D. pseudoobscura*, and *D. virilis*. We screened for predicted binding sites for STAT (Yan, Small, Desplan, Dearolf, & Darnell, 1996), Spalt (Barrio et al., 1996) and for a high-fidelity binding site for Abd-B (Ekker et al., 1994). Putative conserved transcription factor binding sites are listed in Table 2.5.

2.5 SUPPLEMENTAL TABLES

Minimum enhancer	Predicted Binding Site	Extent of conservation	Position	Sequence
Poxn CD2 enhancer	high fidelity AbdB	to *D. vir*	2R:15831429	GCCGTAAA
	STAT	to *D. vir*	2R:15831437	TTCGTGGAA
crb enhancer	weak STAT	to *D. vir*	3R:24297639	TTCGTTTGAA
	weak STAT	to *D. pse* (strong STAT in *D. vir*)	3R:24297680	TTCAGGGGAA
Gef64C CD4 enhancer	weak STAT	to *D. pse*	3L:4696998	TTCCGTGGAA
	weak STAT	to *D. vir*	3L:4697343	TTCTGTTGAA
Cad86C enhancer	Spalt	to *D. pse*	3R:10828636	TTATGTAAT
eya CD6 enhancer	STAT	to *D. vir*	2L:6550780	TTCCGAGAA
ems DS enhancer	high fidelity AbdB	to *D. ana*	3R:13906140	TTTATGGC
	Spalt	to *D. pse*	3R:13905120	TTATGAAAT
	STAT	to *D. vir*	3R:13905989	TTCTCGGAA
	STAT	to *D. vir*	3R:13906567	TTCCTGGAA

Table 2.1. Putative Conserved STAT, Spalt and High Fidelity Abd-B Binding Sites in Minimal Enhancers Shared Between Posterior Spiracle and Posterior Lobe Networks.

Predicted binding sites in minimal co-opted enhancers of *D.melanogaster* were compared to the orthologous regions from *D. simulans*, *D. yakuba*, *D. biarmipes*, *D. ananassae* (*D. ana*), *D. pseudoobscura* (*D. pse*) and *D. virilis* (*D. vir*), only sites conserved to *D. ana*, *D. pse*, and *D. vir* are listed. Consensus binding sites are as follows, STAT: TTCNNNGAA, weak STAT: TTCNNNNGAA, Spalt: TTATGWAMT, high fidelity Abd-B: TTTAYGGC.

Gene	Species	Forward Primer	Reverse Primer
poxn	*D. melanogaster*	ACCGTGGTGAAGAAGGATCATCC	taatacgactcactataggCAGATCAAAACTGGGTCAGTGG
poxn	*D. biarmipes*	atttaggtgacactatagaGAGGAGAACAGCGGCATGTTG	taatacgactcactataggGATTCCACAGCCAGTGCTTGTG
poxn	*D. ananassae*	atttaggtgacactatagaTTCCTTACTAGATTTTCACTGTTC	taatacgactcactataggTCGATGGAGCTCTCCGAC
poxn	*D. pseudoobscura*	atttaggtgacactatagaTTCCTTACTAGATTTTCACTGTTC	taatacgactcactataggCGCTTTGATGGATGTCGTCGTGG
ems	*D. melanogaster*	CGACGAACTGATCGACATGGAGATG	taatacgactcactataggTATAGTTGGTGGTGTCTAGCCTAGG
upd	*D. melanogaster*	TTCTAGTCACATAAGAGCAACCGC	taatacgactcactatagggagaTCAAGCACTATATCACAGAT
upd	*D. biarmipes*	TTCTAGTCACATAAGAGCAACCGC	taatacgactcactatagggagaTCAAGCACTATATCACAGAT
upd	*D. ananassae*	GTAGCTTAAGTAAATTATTTGATTG	taatacgactcactatagggagaGCGGTTGCTCTTATGTGACTAGAA
Cad86C	*D. melanogaster*	ACAACAACGGCACGTTCGAGATCAG	taatacgactcactataggCATCACTTCGCGATCGAAGCCATGC
Cad86C	*D. biarmipes*	ACAGCCAAAGACGAYCTTCATC	taatacgactcactataggTDATCTGCTTGCCATCYGGYTGCTC
Cad86C	*D. ananassae*	ACAGCCAAAGACGAYCTTCATC	taatacgactcactataggTDATCTGCTTGCCATCYGGYTGCTC
Cad88C	*D. melanogaster*	TGCCATAGTGCTAACGCTGACTGAC	taatacgactcactataggATCCTCCAGATCCTTTACCTTCACC
Cad96Cb	*D. melanogaster*	CCATTCAGTACACGATAGTCCAGTC	taatacgactcactataggCATCTTCTCGTAGTCGAGTGGCTTG
crb	*D. melanogaster*	GACAACGGCTATAACCACCTGATCG	taatacgactcactataggATCATCGGACACCTCACCAGGTAAC
crb	*D. ananassae*	CAGACGAACCCCTGCCTGAACAATG	taatacgactcactataggACCAAATATGCTTGCCGCRCGATCC
Gef64C	*D. melanogaster*	GAGACGGAGCTCTTGAAGATTCTTC	taatacgactcactataggGAAATCGAAGAGCTCGTAGTTGTGG
Gef64C	*D. biarmipes*	CGCGATTATGACGATGATGACGAGC	taatacgactcactataggTTYATGCGCAGCGCCATTGTGTCC
Gef64C	*D. ananassae*	CGCGATTATGACGATGATGACGAGC	taatacgactcactataggTTYATGCGCAGCGCCATTGTGTCC
eya	*D. melanogaster*	AAGACCACGCCCACGGGYAAGWC	taatacgactcactataggTGACATCGTCGATGTGCACCTGGTC
eya	*D. biarmipes*	AAGACCACGCCCACGGGYAAGWC	taatacgactcactataggTGACATCGTCGATGTGCACCTGGTC
eya	*D. ananassae*	AAGACCACGCCCACGGGYAAGWC	taatacgactcactataggTGACATCGTCGATGTGCACCTGGTC
en	*D. melanogaster*	TSTGCAAGGCGGTCTCSCAGATYGG	taatacgactcactataggTGGTKGTGGATCCCGTCTCSGARCG
en	*D. biarmipes*	TSTGCAAGGCGGTCTCSCAGATYGG	taatacgactcactataggTGGTKGTGGATCCCGTCTCSGARCG
en	*D. ananassae*	TSTGCAAGGCGGTCTCSCAGATYGG	taatacgactcactataggTGGTKGTGGATCCCGTCTCSGARCG
Gal4	*S. cerevisiae*	atttaggtgacactatagaTCACAGTGTGCAATCCCATTACCGC	taatacgactcactataggGGACCGTTGCTACTGTTAGTGAAAG

Table 2.2. Primers for amplifying species-specific mRNA probes.

Lowercase letters represent sequences for the SP6 (Forward primer) and T7 (Reverse primer) RNA polymerases used for probe synthesis.

Construct	Forward Primer	Reverse Primer
mel poxn PLE	TTCCGggcgcgccTCGGTGGCTTAACACGCGCATT	TTGCCcctgcaggATCGCTGATTCCATGGCCCAGT
yak poxn PLE	TTCCGggcgcgccTTAACGGCCAGCACGAGTTTCC	TTGCCcctgcaggCGTTTTGTTCGAGCGAGTGCAG
ana poxn PLE	TTCCGggcgcgccTTAACGGCCAGCACGAGTTTCC	TTGCCcctgcaggCGTTTTGTTCGAGCGAGTGCAG
pse poxn PLE	TTCCGggcgcgccTTAACGGCCAGCACGAGTTTCC	TTGCCcctgcaggCGTTTTGTTCGAGCGAGTGCAG
mel poxn PLE G4	TTCCGggcgcgccTCGGTGGCTTAACACGCGCATT	TTGCCcctgcaggATCGCTGATTCCATGGCCCAGT
pse poxn PLE G4	TTCCGggcgcgccTTAACGGCCAGCACGAGTTTCC	TTGCCcctgcaggCGTTTTGTTCGAGCGAGTGCAG
mel poxn PLE CD1	TTCCGggcgcgccTCGGTGGCTTAACACGCGCATT	TTGCCcctgcaggGTGATTCGATACGATCCGATGC
mel poxn PLE CD2	TTCCGggcgcgccACTGGTCACTGGACATGGCCAT	TTGCCcctgcaggCCTAAGCCTCCCAATAGAGCGA
mel poxn PLE CD3	TTCCGggcgcgccGTTGATCACATTTCAGCCATGC	TTGCCcctgcaggCCATGGGAAACCAGAAGCTGG
mel poxn PLE CD4	TTCCGggcgcgccTCGCTCTATTGGGAGGCTTAGG	TTGCCcctgcaggATCGCTGATTCCATGGCCCAGT
crumbs spiracle enhancer	TTCCGggcgcgccTAAACGCAGTACGTGGGCGTTGCAC	TTGCCcctgcaggTGTTGTCCGCGGCTCAATTGTTTGG
Gef64C Intron enhancer	TTGCCcctgcaggTTGCCcctgcaggCAACCAACCCACTTGTGAAGGACTG	TTGCCcctgcaggCACGATTCTTCTTCGCCGAAGAACG
Gef64C Intron CD1	TTGCCcctgcaggCACGATTCTTCTTCGCCGAAGAACG	TTCCGggcgcgccCTGCAAGAGGGAGCTCGTCTACAAG
Gef64C Intron CD2	TTGCCcctgcaggCCCAATGAGCATAAAGCTAATGAGG	TTCCGggcgcgccGCAATCCTCAGACACTTAGTCACCG
Gef64C Intron CD3	TTGCCcctgcaggGAAGATCTCCGGTCCGAAGATGTCC	TTCCGggcgcgccCGACAGCTTCCAATTCAACGCGCTC
Gef64C Intron CD4	TTGCCcctgcaggTTGCCcctgcaggCAACCAACCCACTTGTGAAGGACTG	TTCCGggcgcgccAACAGGTCAAGTGCCGCTTGTCTAC
engrailed D enhancer	TTCCGggcgcgccGAATTCGACGCTTAACTAATGATGC	TTGCCcctgcaggGAATTCGCTTGGCTCACACTGAAAC
Cad86C intron enhnacer	TTCCGggcgcgccGCGAAGACAGATACCGAGATGGTC	TTGCCcctgcaggATTAAAGACGTGCTGGACGCGGAAG
eyes absent PLE	ccgggcgaattcgccggcgcgccTCCTAATTCCATCCGACTTTAAGC	cggttgcgatcgcttcctgcaggTGACTTGTTAAATGGGTGTTCC
eyes absent PLE CD5	TTGCCcctgcaggGGAAGGTGGTGGTGGGTTTTTAAGG	TTCCGggcgcgccTACATGACAAAGCTGCTGGGGATGC
eyes absent PLE CD6	TTGCCcctgcaggTCGACCCATCATCATCTTGATGAGC	TTCCGggcgcgccAGGGGTTGGGTAGCTTAAGTTGTCG
ana ems DS enhancer	TTGCCcctgcaggGWRTTYGTCCCACTGTGTGACAGWG	TTCCGggcgcgccGAGATGCATATCAACAATTAGGACGC
ems US hole 1	TTCCGggcgcgccGACGCAAGTCATTCGGGATATGGG	TTGCCcctgcaggCGTGTGGAGCTTGTAAATGACTCAG
ems US hole 2	TTCCGggcgcgccATCATTTACGAAGAAAGCGAGCCGG	TTGCCcctgcaggACGGTAGCCGTCTATCAGATCAGTG
ems US hole 3	TTCCGggcgcgccTCTGAAGAGTTCTCGTCAAGCAGGC	TTGCCcctgcaggCTCACACTGTATCGCCTCCGCTTAG
upd US 1	TTGCCcctgcaggTGCTATCACTGTTCCTCCCTGACTAG	TTCCGggcgcgccCTGAGAAATGGGAAACTCACACCTC
upd US 2	TTGCCcctgcaggCTTTCGAGGGCTTGCACAATTGACG	TTCCGggcgcgccCATACGCGTACCACCATACTCACTG
upd DS1	TTCCGggcgcgccTCCTGGCGCCATATCAATTACACTC	TTGCCcctgcaggTCGGATGCAAAGTATGTGCACATGG
upd DS2	TTCCGggcgcgccCTCTTGACCTTTTGCGGCTATTTGG	TTGCCcctgcaggTCCAGTACACATATCTTCGCGTAGG
upd DS3	TTCCGggcgcgccCTTTCGTCGTCAGCT	TTGCCcctgcaggTCATCTCATCTCAG

	CGTCAGTTTG	CTCCAGACACC
upd DS4	TTCCGggcgcgccGTTCACCTTGTTTATGGACTCGCTG	TTGCCcctgcaggAGACAGAGAGAGGGGATCAGAAACC
upd DS5	TTCCGggcgcgccATGCATCAATTAGCTCCCACTGAGC	TTGCCcctgcaggGTAGCGGTAGCAAAAGGCTACTAAC
upd DS6	TTCCGggcgcgccGAGATGCTGTGCCGGTGATTATGAC	TTGCCcctgcaggACCGACATATGACTAAGCCAGCAGC

Table 2.3. Primers used for transgenic constructs.

Lowercase letters represent restriction sites for *AscI* (Forward primers) and *Sbf I* (Reverse primers) used for cloning

Mutation	**Forward Primer**	**Reverse Primer**
mel poxn PLE STAT mutant	CGCCGTCCGAAaaCGTGGAAGCCG	CGGCTTCCACGttTTCGGACGGCGTCCAGGCCC
mel poxn PLE Abd-B mutant	CGAATTCGTGGAAGCCGccgAAAGTCTTCGGGGAGTG	CACTCCCCGAAGACTTTcggCGGCTTCCACGAATTCG
mel *eya* STAT mutant	CTGCAGCTCaaCCGAGAATTTGGTACGAG	CTCGTACCAAATTCTCGGttGAGCTGCAG

Table 2.4. Primers for generating mutant binding site reporters by overlap extension PCR.

Lowercase letters represent altered bases introduced into the mutant construct to disrupt the binding sites.

Reporter Name	Species	51D9, GFP	68A4 (attP2), GFP	68A4 (attP2), Gal4	68E1, Gal4
Poxn posterior lobe enhancer	mel	X (Fig. 3D, 3E, 12B, 12B', 5B, 5B', 5B")			X (Fig S3I')
Poxn posterior lobe enhancer	yak	X (Fig. 5G, 5G')			X
Poxn posterior lobe enhancer	ana	X (Fig. 3H, 3I, 5H, 5H')			X
Poxn posterior lobe enhancer	pse	X (Fig. 5I, 5I')			X (Fig S3M')
Poxn posterior lobe enhancer CD1	mel	X (Fig. 5C, 4C', 4C")			
Poxn posterior lobe enhancer CD2	mel	X (Fig. 5D, 5D', 5D")			
Poxn posterior lobe enhancer CD3	mel	X (Fig. 5E, 5E', 5E")			
Poxn posterior lobe enhancer CD4	mel	X (Fig. 5F, 5F', 5F")			
Poxn enhancer STAT mutant	mel	X (Fig. 12C, 12C', 5J)			
Poxn enhancer AbdB mutant	mel	X (Fig. 12D, 12D', 5K)			
crb enhancer	mel	X	X (Fig. 10G, 10G')	48944	
crb US	mel	X			
crb Rubin Gal4 line with no activity	mel			48851	
crb Rubin Gal4 line with no activity	mel			48877	
crb Rubin Gal4 line with no activity	mel			48918	
crb Rubin Gal4 line with no activity	mel			45455	
en D enhancer	mel	X (Fig. 10H, 10H')			
Gef64C Intron + Exon 1	mel	X	X		
Gef64C IE CD1	mel	X			
Gef64C IE CD2	mel	X			
Gef64C IE CD3	mel	X			
Gef64C IE CD4	mel	X (Fig. 10I, 10I')			
Cad86C enhancer	mel		X (Fig. 10J,10J')		
eya enhancer	mel	X	X (Fig. 12E,12E')	48893 (Fig. 10K,10K')	
eya enhancer STAT mutant	mel		X (Fig. 12F,12F')		
eya enhancer CD5	sech	X (Fig. 11C, 11C')			
eya enhancer CD6	sech	X (Fig. 11D, 11D')			
eya Janellia Lines	mel			48881	
eya Rubin Gal4 line with no activity	mel			47890	

eya Rubin Gal4 line with no activity	mel			48891
eya Rubin Gal4 line with no activity	mel			45837
eya Rubin Gal4 line with no activity	mel			48897
eya Rubin Gal4 line with no activity	mel			49292
eya Rubin Gal4 line with no activity	mel			48930
ems US enhancer	mel			47890 (Fig. 10L, 10L', 10O, 10O')
ems DS enhancer	mel			40523 (Fig. 10M, 10M', 10P, 10P', 11A, 11A')
ems DS enhancer	ana		X (Fig. 11B, 11B')	
ems DS enhancer CD1+2	mel	X		
ems US hole 1	mel	X		
ems US hole 2	mel	X		
ems US hole 3	mel	X		
ems Rubin Gal4 line with no activity	mel			48387
ems Rubin Gal4 line with no activity	mel			46846
ems Rubin Gal4 line with no activity	mel			40510
ems Rubin Gal4 line with no activity	mel			41318
ems Rubin Gal4 line with no activity	mel			46847
ems Rubin Gal4 line with no activity	mel			46848
ems Rubin Gal4 line with no activity	mel			47828
ems Rubin Gal4 line with no activity	mel			48397
ems Rubin Gal4 line with no activity	mel			40522
ems Rubin Gal4 line with no activity	mel			49423
ems Rubin Gal4 line with no activity	mel			47981
ems Rubin Gal4 line with no activity	mel			49424
ems Rubin Gal4 line with no activity	mel			46856
ems Rubin Gal4 line with no activity	mel			46857
ems Rubin Gal4 line with no activity	mel			46858
ems Rubin Gal4 line with no activity	mel			40528
ems Rubin Gal4 line	mel			46865

with no activity			
ems Rubin Gal4 line with no activity	mel		40529
upd enhancer DS1	mel	X	
upd enhancer DS2	mel	X	
upd enhancer DS3	mel	X	
upd enhancer DS4	mel	X	
upd enhancer DS5	mel	X	
upd enhancerDS6	mel	X (Fig. 13E, 13F)	
upd enhancer US1	mel	X	
upd enhancer US2	mel	X	

Table 2.5. Transgenic lines analyzed.

For each construct generated, the species from which it was cloned is listed, as well as the insertion site, and type of reporter (GFP or GAL4). For GAL4 constructs from the Rubin collection, the Bloomington Drosophila Stock Center stock number is listed. For each transgenic line that was presented in a figure, the panel number is provided.

3.0 A CASCADE OF INTERCELLULAR SIGNALING PATHWAYS CONTRIBUTE TO THE ORIGINATION OF THE POSTERIOR LOBE

The following collaborators contributed data presented in this chapter: Chas Elliot and Mark Rebeiz performed the initial antibody screen of key *Drosophila* signaling ligands and transcription factors. Kristen Gardner performed the *in situ* hybridization to detect *Drop* mRNA (Figure 3.8). Mark Rebeiz perturbed Notch signaling to assay its effects on posterior lobe development (Figure 3.3). Chas Elliot performed antibody staining on lobed and non-lobed genitalia to compare patterns of Delta (Figure 3.4). Winslow Johnson cloned the downstream regulatory fragments of *Delta* into our transgenic reporter system (Figure 3.6). Winslow Johnson and Stephanie Day screened and homozygosed most of the transgenic insertions in this chapter. Natalie Dall analyzed non-lobed species' *Delta* posterior lobe reporters (Figure 3.7). Sarah Smith performed the analysis of the *unpaired* regulatory region.

3.1 INTRODUCTION

The types of changes underlying the evolution of novel morphologies is a matter of open debate in the field of evolutionary developmental biology. Complex gene regulatory networks (GRN's) are required to orchestrate the development of many morphologies, but it is not known whether they evolve in small increments or through large, transformative events (Frazzetta, 2012).

Network co-option, or the redeployment of a pre-existing GRN to a new location, has been presented as a mechanism to rapidly evolve novel morphologies (Monteiro & Podlaha, 2009; Rebeiz et al., 2015). Although this mechanism has been found to contribute to the evolution of evolutionary novelties such as the sea urchin larval skeleton (Gao & Davidson, 2008) the beetle horn (Moczek & Nagy, 2005), and the turtle's shell (Kuraku et al., 2005), the precise molecular changes responsible for the origination of these co-option events has been harder to pinpoint. Identifying the mechanisms underlying network co-option would help to address important questions about the process of morphological evolution.

Signaling pathways are potent sources of pattern during development, allowing one or a few cells to organize the behavior of surrounding tissues during embryogenesis, cell type differentiation and organ formation (Perrimon, Pitsouli, & Shilo, 2012). One such pathway, acting through the Notch receptor, is an important mechanism for contact-dependent signaling, in which one group of cells can instruct their neighbors (Guruharsha, Kankel, & Artavanis-tsakonas, 2012). This pathway is initiated by the binding of a Notch ligand (either *Delta* or *Serrate* in *Drosophila*) to the Notch receptor of a neighboring cell, causing the Notch receptor to self-cleave and release its intracellular domain (Figure 3.1). The intracellular domain of Notch localizes to the nucleus and binds to the transcription factor *Suppressor of Hairless* (*Su(H)*), displacing a co-repressor complex (Maier, 2006), and converting it to a transcriptional activator, thus activating Notch signaling (Figure 3.1). In contrast to this *trans* activation of neighboring cells, Notch ligands can produce a *cis* inhibition of their own cell's Notch receptors, giving directionality to the signal (Guruharsha et al., 2012).

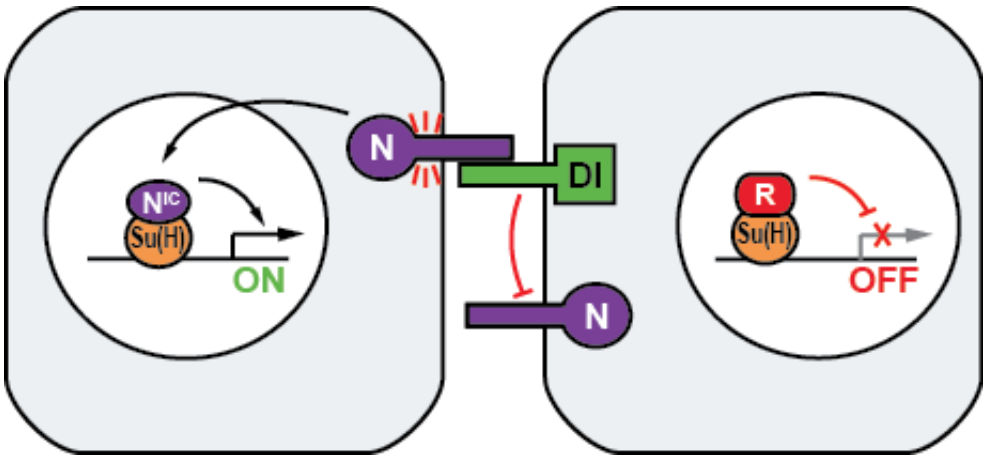

Figure 3.1. Notch Signaling Pathway.

A schematic of the Notch signaling pathway. (N: Notch, N^{IC}: intracellular domain of Notch, Dl: Delta, Su(H): Suppressor of Hairless, R: repressor).

The posterior lobe is a recently evolved morphological novelty in the *Drosophila melanogaster* clade (Kopp & True, 2002). This lobe inserts between the 6th and 7th abdominal segments during copulation and is necessary for copulation (Frazee & Masly, 2015; Kamimura, 2010). We recently discovered that a portion of the posterior spiracle network was co-opted into the posterior lobe during its evolution, but the causative mechanism underlying this co-option remains unclear (Glassford et al., 2015). Given the widespread importance of signaling pathways to many developmental processes, we sought to identify pathways that may contribute to the development and evolution of the posterior lobe, and may have therefore played a role in this structure's origination. We identified a posterior lobe associated pattern of the Notch ligand Delta that is necessary for deployment of the posterior lobe's GRN and morphogenesis. Screening the genomic region surrounding the *D. melanogaster Delta* gene, I identified an enhancer element that recapitulates its lobe-patterning activity. By comparing the activity of this

enhancer element from lobed and non-lobed species in multiple *trans* regulatory landscapes, I conclude that the species-specific pattern of *Delta* expression arose through changes upstream of this enhancer. To identify *trans*-factors whose alterations can account for the observed changes in Delta expression, I identified several likely regulators of *Delta*, including a regulator that is specifically necessary for the lobe-associated expansion of *Delta* activity.

3.2 RESULTS

3.2.1 A Species Specific Expansion of Delta Expression is Necessary for Development of the Posterior Lobe

In a screen of key *Drosophila* signaling pathway ligands, we identified the Notch ligand Delta as a candidate upstream factor whose expression precedes posterior lobe development. A developmental timecourse of Delta antibody staining during genital development revealed expression in several locations, including a pattern near the prospective lobe tissue (Figure 3.2A, arrow). From just before the initial emergence of the posterior lobe at 28h to midway through the lobe's development at 40h, *Delta* activity can be seen extending from its initial zone of expression (Figure 3.2B-F). From 40h to 48h this pattern retracts toward the anal plate (Figure 3.2F-H) at the same time as the posterior lobe emerges from the adjacent epithelium (Glassford et al., 2015). The patterned expression of Delta in a dynamic profile that correlates with the timing and positioning of the posterior lobe suggested that Notch-Delta signaling may play a key role in the development of this novel structure.

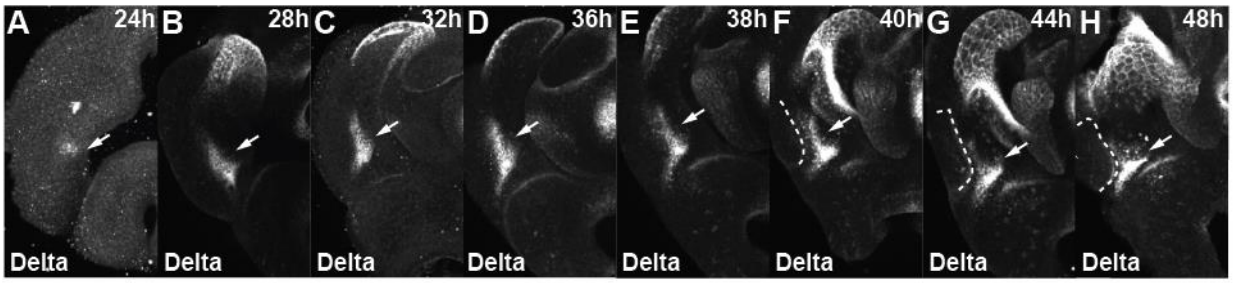

Figure 3.2. Dynamic Expansion and Contraction of Delta Expression During Posterior Lobe Development.

(A-H) Pupal male genital samples stained with antibody targeting Delta at timed intervals following puparium formation. Arrows highlight *Delta* expression associated with the posterior lobe. Delta is first active in a signaling center (A), and then expands along the furrow between the cleaving lateral plate and clasper prior to posterior lobe development (B-F). Delta recedes to the signaling center as the posterior lobe begins to develop (F-H). Times are listed in hours after puparium formation (h).

To investigate the possible role for Notch-Delta signaling during posterior lobe development, we employed the GAL4-UAS system, using a genital-specific *Poxn* driver to reduce and increase this pathway's activity (Boll & Noll, 2002). The selected *Poxn* driver is active in a broad zone that encompasses the posterior lobe and surrounding tissues during development (not shown). Compared to a wild-type control (Figure 3.3A), the RNAi knockdown of *Delta* almost completely abolished posterior lobe development when driven by the *Poxn* driver (Figure 3.3B). In contrast, over-activation of the pathway by expressing the constitutively active Notch intracellular domain under control of the *Poxn-GAL4* driver greatly increased the size of the posterior lobe (Figure 3.1C). These two experiments show that Notch signaling is both necessary for posterior lobe development, and able to modulate the size of the posterior lobe.

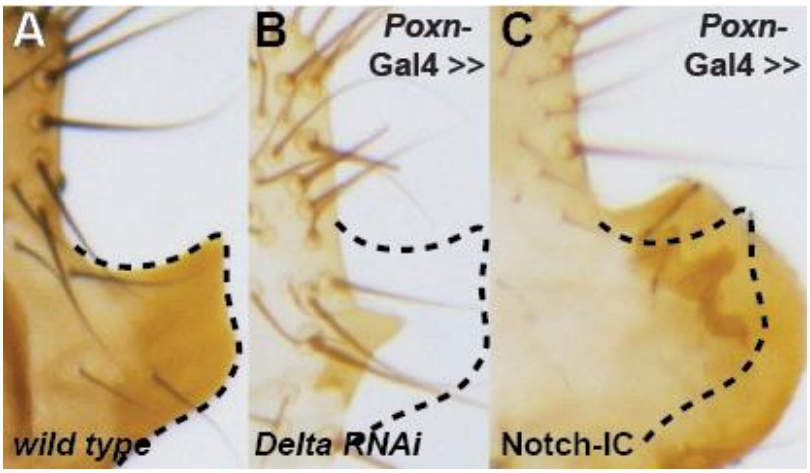

Figure 3.3. Notch Signaling Modulates Posterior Lobe Development.

(A-C) Expression of RNA hairpins targeting the Notch ligand *Delta* (B) inhibit posterior lobe development in comparison to a wild type *D. melanogaster* (A), while expression of the intracellular domain of Notch (Notch-IC) increases the size of the posterior lobe (C).

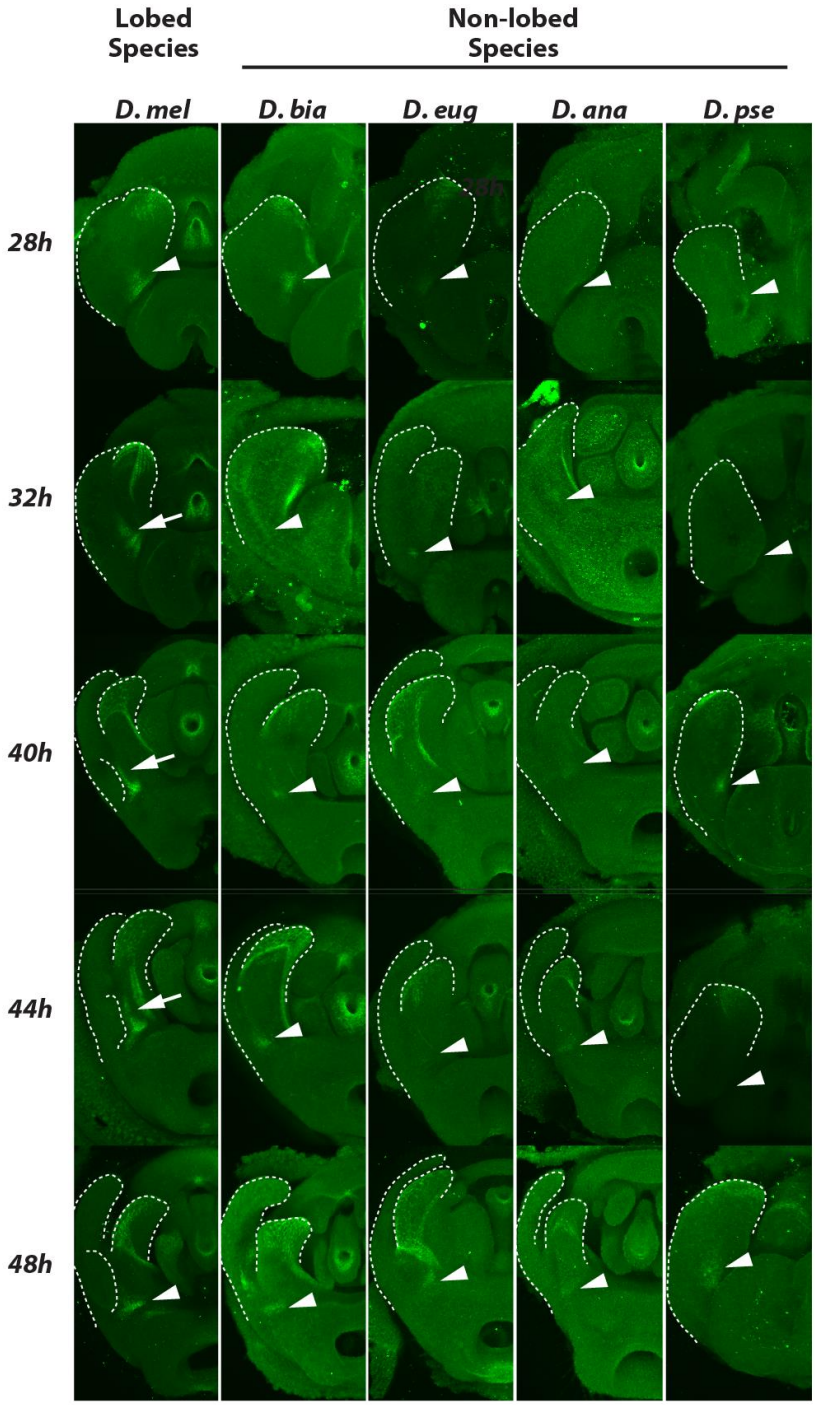

Figure 3.4. Comparative Analysis of Delta Expression Reveals a Species-specific Expansion Unique to Lobed Species.

Timed pupal male genitals from lobed (*D. melanogaster*) and non-lobed species (*D. biarmipes*, *D. eugracilis*, *D. ananassae*, *D. pseudoobscura*) stained with a cross-reactive antibody specific to Delta protein at 4

hour intervals. Delta's expression pattern expands from a signaling center prior to posterior lobe development (28h-40h) in lobed species, but does not in non-lobed species. Arrows mark expanded lobe-associated pattern of *Delta*. Arrowheads mark non-expanded *Delta* pattern.

The drastic posterior lobe phenotypes that resulted from manipulating the Notch pathway indicated that this signaling pathway plays a key role during the development of the posterior lobe. We were next curious whether the lobe associated patterns of Delta expression were specific to species that develop this genital structure. Using a polyclonal Delta antibody that broadly cross-reacts with several *Drosophila* species, we compared its expression with multiple species that lack posterior lobes over a broad swath of genital development (Figure 3.4). Each species initially expressed Delta in a pattern that resembled the initial expression of Delta in *D. melanogaster* at 28h before it expands into its lobe-associated pattern. While the pattern of Delta accumulation intensifies and expands spatially in *D. melanogaster*, in each non-lobed species, this pattern did not expand (Figure 3.4). This phenomenon is similar to the activity of the Jak/STAT ligand *upd* in lobed and non-lobed species (Glassford et al., 2015), in which *upd* expression is extended during posterior lobe development in lobed species but in non-lobed species *upd* expression disappears during the analogous period in time. Since both pathways exhibit similar spatiotemporal dynamics across many species, it is possible that they may regulate one another.

To resolve the genetic relationship between Notch signaling and Jak/STAT signaling, we expressed transgenic RNAi hairpins targeting *Delta* or a control gene (*mCherry*) in the context of a transgenic reporter containing ten multimerized STAT92E binding sites (10XSTATGFP) which provides a readout of JAK/STAT signaling (Bach et al., 2007). RNAi knockdown of *Delta* drastically reduced the 10XSTATGFP reporter signal in comparison to the control during late

posterior lobe development (48h) (Figure 3.5B-C). This result suggests that Notch signaling is upstream of Jak/STAT signaling at this stage. To uncover which aspects of *Delta* expression were responsible for activating the JAK/STAT pathway, we examined Delta expression in the *Delta* RNAi knockdown animals. This experiment revealed that only the expanded lobe-associated pattern of Delta is disrupted (Figure 3.5A). Collectively, these results indicate that the expanded posterior lobe associated pattern of Delta is necessary for the late, high levels of Jak/STAT signaling during posterior lobe development (Figure 3.5B). The small pattern of Delta that remains upon RNAi knockdown (Figure 3.5A, arrow) is insufficient to induce the upregulation of Jak/STAT signaling late in genital development and resembles the pattern present in non-lobed species (Figure 3.4).

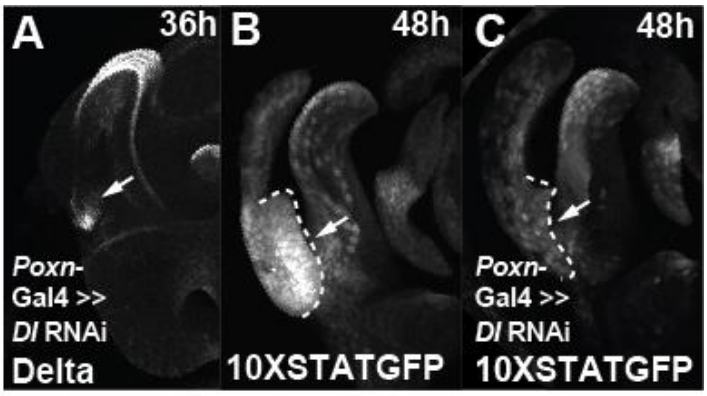

Figure 3.5. The Posterior Lobe Associated Pattern of Delta is Necessary for the Upregulation of Jak/STAT Signaling Within the Developing Posterior Lobe.

(A) RNA hairpins targeting *Delta* inhibit the lobe-associated pattern of Delta expression, but do not inhibit *Delta* signaling center activity (arrow). A reporter containing 10 concatenated *STAT92E* binding sites in the presence of RNA hairpins targeting *Delta* shows a reduction in activity in comparison to a reporter in a wild-type background (B-C).

3.2.2 The *Delta* Posterior Lobe Enhancer Is Not Novel to Lobed Species

We next sought to discover whether the expanded pattern of *Delta* unique to posterior lobed species is caused by a *cis* regulatory change at *Delta*, a task which necessitated the identification of its posterior lobe specific enhancer. A screen for genital enhancers at *Delta* covered ~105kb of non-coding DNA surrounding the locus, which included overlapping ~5kb transgenic reporters from the Janelia Gal4 collection (Pfeiffer et al., 2008), as well as downstream and intronic fragments cloned into our transgenic reporter system (Figure 3.6A). We collected timed genital samples and used fluorescent confocal microscopy to screen for patterns recapitulating the endogenous Delta expression patterns. Among 31 constructs, two reporters, downstream tiles (DS) 3 and 4, reported activity in a pattern that recapitulated the expanded Delta pattern at 36hAPF (Figure 3.6B-C). Hypothesizing that there was a single posterior lobe enhancer contained within the overlap of these two tiles, we cloned a reporter for the ~2kb region of overlap between tiles 3 and 4 into our transgenic reporter system, which also drove the same pattern (Figure 3.6D). This reporter also drove expression in a portion of the hypandrium (Figure 3.5D arrowhead) but not in the clasper when inserted into a distinct landing site, 68A3. These results indicate that enhancers for the posterior lobe and hypandrium patterns reside within this fragment of DNA.

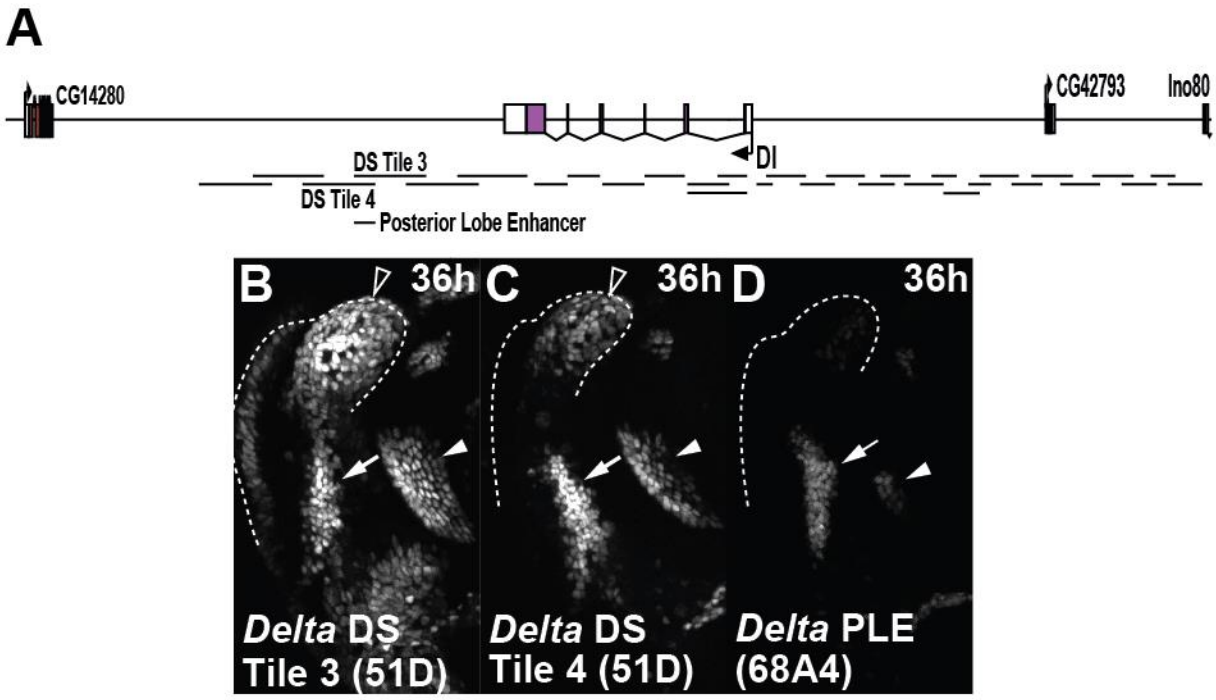

Figure 3.6. A Survey of the *Delta* Regulatory Region Identifies a Posterior Lobe Enhancer.

(A) A schematic of the *Delta* gene and its surrounding regulatory region, black lines represent regions of regulatory DNA screened for genital activity in a transgenic reporter system (B-D) *Delta* downstream (DS) tile 3 (B) and tile 4 (C) reporters exhibit the expanded pattern of *Delta* (arrows) and a pattern in the hypandrium (filled arrowheads). A reporter covering the overlap of tiles 3 and 4, named *Delta* posterior lobe enhancer (PLE), reports the expanded *Delta* pattern (D, arrow) and hypandrium pattern (arrowhead). Transgenic reporters inserted into the 51D landing site (B-C) exhibit ectopic activity in the clasper (arrowhead outline) while a reporter in 68A4 (D) does not.

Given that *Delta* is necessary for the deployment of the posterior lobe program, we next sought to assess the possibility that its expanded pattern arose as a result of a modification to its enhancer in posterior lobed species. To assess this possibility, the orthologous regulatory region of *Delta* was cloned from several non-lobed species and the extent of their patterns of expression were compared within the same transgenic *D. melanogaster* context. The *D. biarmipes* and *D. ananassae Delta* posterior lobe enhancer reporters drove expression that overlapped, and

extended beyond the endogenous pattern of Delta recapitulated by the *D. melanogaster* enhancer. This result indicates that if *cis* changes have occurred at *Delta*, they were not necessary for its expanded pattern (Figure 3.7A-C).

Several other possible mechanisms could explain the discrepancy between the non-lobed species' *Delta* posterior lobe enhancer reporters and the much weaker endogenous deployment of Delta in non-lobed species. The *Delta* posterior lobe enhancer could be auto-regulatory, and could be over-activated by the high levels of Delta deployment in the *D. melanogaster Delta* activity. Also, in non-lobed species the *Delta* posterior lobe enhancer may be repressed by a silencer that is not contained in the 2kb fragment cloned in our reporters. Finally, the *trans* landscapes in which *Delta* is expressed might be too different between lobed and non-lobed species to properly compare differences in spatial patterning. To confirm that changes at *Delta* are not sufficient to expand *Delta* expression, I assembled a construct that drives expression of *D. melanogaster Delta* cDNA using the *D. melanogaster Delta* posterior lobe enhancer within a PiggyBac transposase vector backbone, and inserted it transgenically into the *D. ananassae* genome (Figure 3.7D). To observe only transgenic *D. melanogaster Delta* expression without capturing endogenous *D. ananassae Delta* signal, I performed antibody staining on timed transgenic *D. ananassae* genital samples with the *D. melanogaster* clade-specific monoclonal antibody that will only recognize Delta produced by the transgene. The *D. melanogaster* enhancer drove a pattern of Delta expression that strongly resembled the ancestral pattern endogenous to the background in which it was inserted (Figure 3.7E, arrow), no expansion of *Delta* in the lateral plate was observed across several points of genital development (not shown). As an internal control, the lobed-species specific hypandrium pattern is driven by this construct as well, confirming that the transgenic insertion is able to drive proper expression in a

neighboring tissue (Figure 3.7E, arrowhead). In sum, these data suggest that the *D. melanogaster* posterior lobe enhancer of *Delta* is unable to drive its species-specific pattern without other *trans* regulatory changes in the genome.

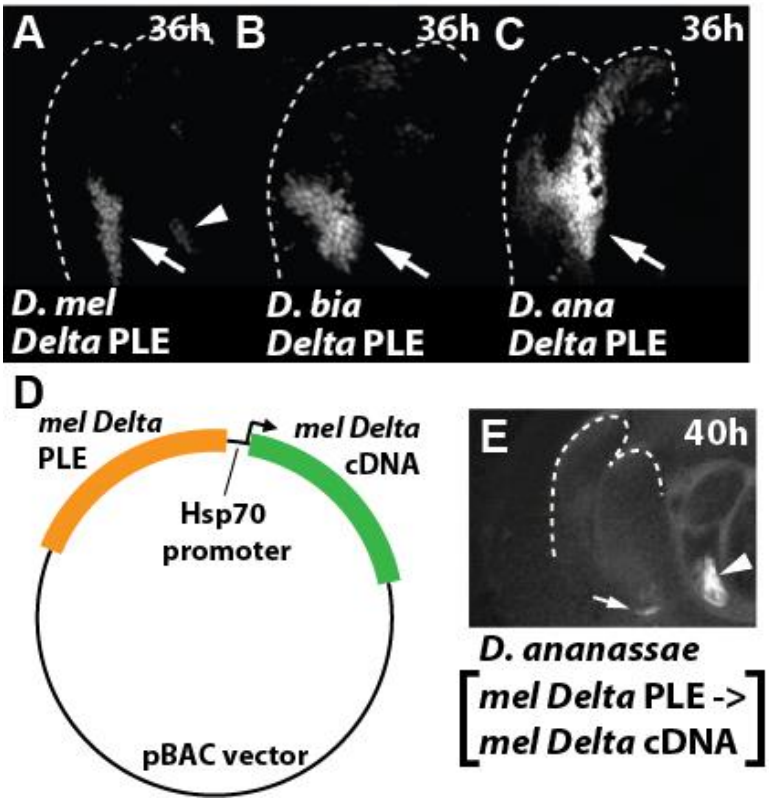

Figure 3.7. *trans* Regulatory Change is Necessary for the Expanded Expression of Delta in Association with the Posterior Lobe.

(A-C) Orthologous transgenic reporters of the *Delta* posterior lobe enhancer (PLE) from lobed *D. melanogaster* (A), and non-lobed *D. biarmipes* (B) and *D. ananassae* (C) report an expanded pattern in transgenic *D. melanogaster* flies. (D) Schematic of a piggyBAC (pBAC) transgenic vector construct containing the *D. melanogaster* PLE (orange) driving *D. melanogaster Delta* cDNA (green) fused to an Hsp70 promoter. (E) *D. melanogaster*-specific antibody staining reveals an *annanassae* specific expression pattern of Delta when driven by the *melanogaster* enhancer transgenically placed into the *D. ananassae* genome. Arrows mark *Delta* posterior lobe associated pattern of *Delta*. Arrowheads highlight the hypandrium associated pattern.

3.2.3 Identification of *trans* regulators of *Delta*

The findings above provided strong evidence that *trans*-regulatory changes upstream of Delta account for its expanded expression in species that develop a posterior lobe., Hence, I next sought to elucidate what factor(s) regulate the *Delta* enhancer during its initial establishment. To identify sex-specific gene activity in the genitalia, a microarray-based screen was used to identify genes differentially expressed between male and female 3rd instar larval genital imaginal discs, 6h pupal genitalia and 20h pupal genitalia (Chatterjee, Uppendahl, Chowdhury, Ip, & Siegal, 2011). Of the several genes identified, only *Drop* was found to be specifically necessary for external male genital structures (Chatterjee et al., 2011). Since the posterior lobe was one of the external male genital structures found to require *Drop* activity, we performed an *in situ* hybridization for *Drop* in early male genitalia to detect *Drop* mRNA localization just prior to the development of the posterior lobe. At 32h, *Drop* was found to be expressed between the presumptive lateral plate and clasper, a location overlapping the posterior lobe associated pattern of *Delta* (Figure 3.8A).

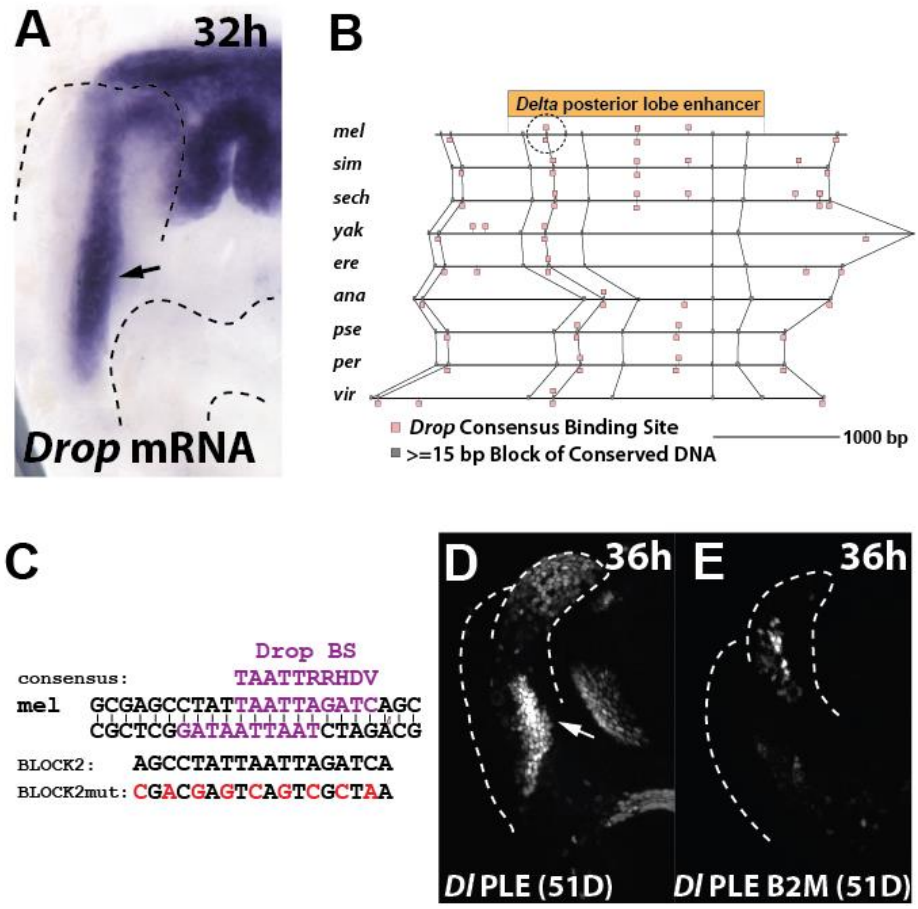

Figure 3.8. Sequence Containing Predicted Binding Sites for *Drop* is Required for the Activity of the *Delta* Posterior Lobe Enhancer.

(A) *in situ* hybridization detects *Drop* mRNA in 32h male pupal genitalia. (B) Alignment of orthologous sequences from several lobed and non-lobed species covering the *Delta* posterior lobe enhancer region (orange bar). (Pink boxes) Predicted Drop binding sites (Gray boxes) Blocks of perfectly conserved sequence ≥15 base pairs in length between listed *Drosophila* species. (C) Drop consensus binding site and predicted Drop binding sites within conserved block "2" are highlighted purple. The block "2" mutations introduced into the *Delta* posterior lobe enhancer reporter are highlighted red. (D-E) Altering every other base of block 2 in the context of the *Delta* posterior lobe enhancer (E) completely ablates posterior lobe associated expression in comparison to the control (D, arrow).

Drop encodes a transcription factor, and it is possible that *Drop* may directly regulate the activity of *Delta* through the *Delta* posterior lobe enhancer. To resolve this possibility, I examined the region orthologous to the *Delta* posterior lobe enhancer from several lobed and non-lobed species for consensus Drop binding sites, and discovered two predicted sites within an 18bp sequence conserved in all assayed *Drosophila* species (Figure 3.8B-C). To test the contribution of this sequence to the activity of *Delta*'s posterior lobe enhancer, we scrambled the 18bp block of conserved DNA by introducing a non-complementary transversion to every other base in the context of the *Delta* PLE (Figure 3.8C). The resulting *Delta* posterior lobe enhancer completely lacked posterior lobe activity in comparison to the unaltered reporter (Figure 3.8D-E). This result is consistent with the hypothesis that *Drop* directly regulates *Delta* through its posterior lobe enhancer.

Flies mutant for *Poxn* exhibit reduced adult posterior lobes, claspers, and penises among other defects, but are viable through pupal development (Boll & Noll, 2002). Earlier work by our lab found *Poxn* to be active in a wave of expression during the cleavage of the clasper from the lateral plate (Glassford et al., 2015). This pattern overlaps the lobe associated pattern of Delta both spatially and temporally, and therefore represents a candidate regulator of the posterior lobe activity of *Delta*. To assess the contribution of *Poxn* to the genital activities of *Delta*, we assessed Delta expression in flies mutant for *Poxn* at several times during pupal. In the absence of *Poxn*, *Delta* maintains expression in the initial pattern near the anal plate, but never expands into the lateral plate tissue (Figure 3.9B-E). This indicates that *Poxn* is necessary for the lobe-associated expansion of *Delta*, but is dispensible for the ancestral non-extended pattern of *Delta*. To evaluate the possibility that *Poxn* directly binds to the *Delta* posterior lobe enhancer, we aligned the orthologous regulatory regions from several lobed and non-lobed species and

screened for sequences predicted to bind *Poxn* in a bacterial 1-hybrid assay (Noyes et al., 2008). One predicted strong binding site was identified, and was conserved between the *D. melanogaster* and *D. yakuba* clades (Figure 3.10A). Alteration of two bases in the predicted *Poxn* binding site within the context of the *Delta* posterior lobe reporter did not disrupt reporter expression, suggesting *Poxn* may instead indirectly regulate the posterior lobe associated expression of Delta (Figure 3.10B-D).

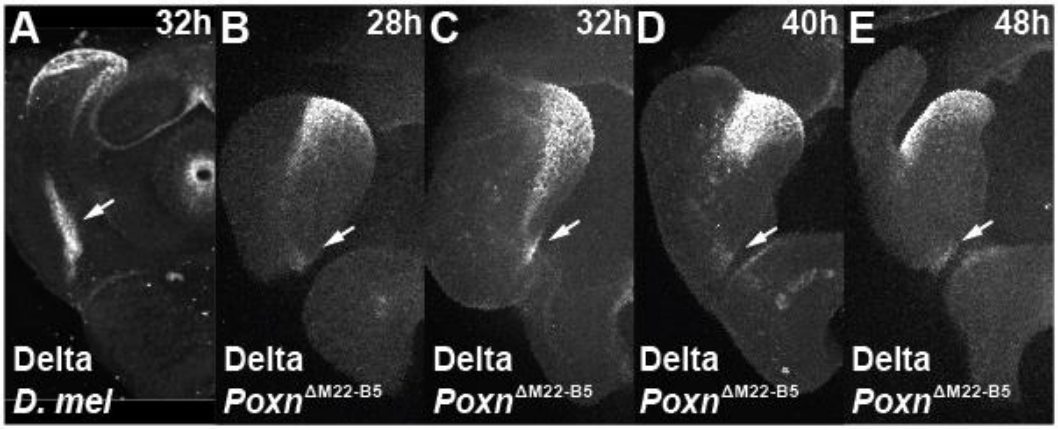

Figure 3.9. *Poxn* is necessary for the expansion of *Delta* from the signaling center.

(A-E) Antibody staining for Delta reveals flies mutant for *Poxn* (B-E) do not exhibit an expanded pattern of *Delta* from its signaling center that occurs in wild type flies (A). Arrows mark the pattern of *Delta* associated with the posterior lobe.

Figure 3.10. Mutation of a predicted binding site for *Poxn* does not disrupt *Delta*'s posterior lobe enhancer.

(A) Alignment of orthologous sequences from several lobed and non-lobed species covering the *Delta* posterior lobe enhancer region (orange bar). (B) Alignment of a consensus binding site for Poxn (Noyes et al., 2008), the predicted Poxn binding site in *D. melanogaster*, and the mutation introduced into the *Delta* posterior lobe enhancer reporter. (C-D) Introducing the *Poxn* binding site mutations into the *Delta* posterior lobe enhancer (D) does not disrupt posterior lobe associated expression in comparison to the control (C).

3.2.4 The *Delta* Posterior Lobe Enhancer is Active in the Longitudinal Visceral Mesoderm

In an earlier chapter, direct regulators of the *Poxn* and *Eya* posterior lobe enhancers were easily identified because they had been co-opted from the well-studied posterior spiracle network, providing several candidate *trans* regulators. Although our experiments indicate that the *Delta* posterior lobe enhancer was ancestrally active in the genital signaling center before the origination of the posterior lobe, it is possible that it also shares inputs with other tissues from which its derived activities were co-opted. To assess this possibility, I screened the *Delta* PLE reporter for expression in additional developing tissues. Since several other posterior lobe enhancers had been co-opted from the posterior spiracle network I first looked for reporter expression during embryogenesis, but I did not observe expression within any part of the posterior spiracle (Figure 3.11A, arrowhead). However, reporter activity was visible in another embryonic tissue, the longitudinal visceral mesoderm (LVM, also known as the caudal visceral mesoderm) (Figure 3.11A-B, arrows). Before embryonic stage 13, the LVM migrates anteriorly alongside trunk visceral mesoderm (TVM), before fusing with cells from the TVM to form the multinucleated longitudinal muscles of the embryonic midgut (Lee, Lee, Zaffran, & Frasch, 2005). The LVM can be visualized using a transgenic line containing the LVM-specific factor *HLH54F* recombineered with a fluorescent protein (Figure 3.11C-D) (Ismat et al., 2010). *Delta*

posterior lobe enhancer reporter activity was not observed before the migration of the LVM at embryonic stage 13 (data not shown).

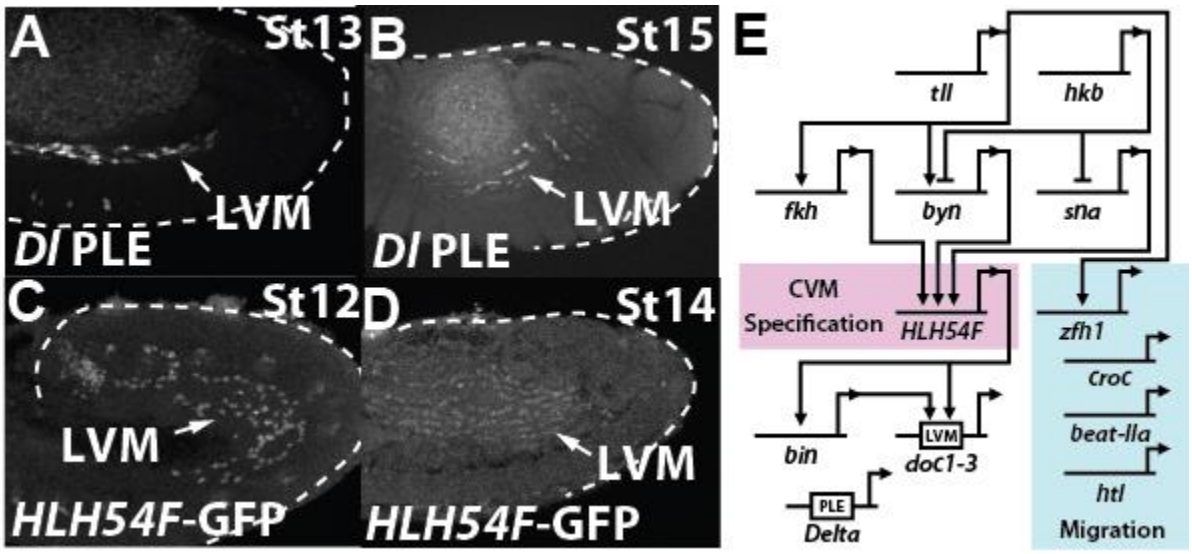

Figure 3.11. *Delta* posterior lobe enhancer (PLE) is expressed in the LVM.

(A-B) *Delta* PLE is expressed during the migration of the longitudinal visceral muscle (LVM) (A) and in the longitudinal muscle syncytia (LVS) (B). (C-D) Fluorescently-tagged LVM marker gene *HLH54F* is visible during LVM migration and in the LVS (E) Several terminal and ventral embryonic patterning genes specify the CVM by regulating the expression of *HLH54F* (pink). Several CVM genes do not require *HLH54F* activity, but are required for CVM migration (blue). Modified from (Ismat et al., 2010).

Several genes are known to regulate the specification and morphogenesis of the LVM, but fewer are expressed during or just before LVM migration. The bHLH transcription factor *HLH54F* is specified early in embryonic development by several terminal patterning genes and *snail* (Figure 3.11E). After specification, *HLH54F* is continuously expressed in the LVM primordium during gastrulation (embryonic stage 5) and is maintained to at least the formation of the longitudinal muscle syncytia (embryonic stage 15) (Ismat et al., 2010). *binou* (*bin*) is expressed in all visceral muscle tissue (Zaffran, Küchler, Lee, & Frasch, 2001), but requires

HLH54F for its activity in the LVM (Ismat et al., 2010). One or more of the *Dorsocross* (*Doc*) complex genes are expressed in the LVM between embryonic stages 10 and 12 (Reim, Lee, & Frasch, 2003); this activity requires the presence of *HLH54F* and likely *bin*, which cooperatively regulate an LVM-specific enhancer in the *doc* locus (Ismat et al., 2010). Additional genes expressed during LVM migration are *crocodile* (*croc*) (Hackerl et al., 1995), *beat-IIa* (Tomancak et al., 2002), Zinc-finger transcription factor *ZfhI* (Broihier, Moore, Doren, Newman, & Lehmann, 1998), and *heartless* (*htl*) (Beiman, Shilo, & Volk, 1996).

3.2.5 *Doc2* is a Candidate Regulator of *Delta*'s Posterior Lobe Enhancer

Any LVM-specific transcription factor that is present before or during the activation of the *Delta* posterior lobe enhancer in the LVM may be a regulator of *Delta*, and may potentially lie upstream of this element in the genitalia. I therefore screened several LVM genes for expression during genital development, beginning with *HLH54F*. A probe for *HLH54F* mRNA (Figure 3.12), and a transgenic line containing *HLH54F* recombineered with a fluorescent tag (not shown), did not express a detectible signal in the genitalia, suggesting that *HLH54F* is not active prior to posterior lobe development. While *in situ* hybridization for the mRNA transcripts of *bin*, *zfh1*, and *fkh* revealed known patterns of their expression during embryonic development, these probes were not expressed prior to lobe development (not shown).

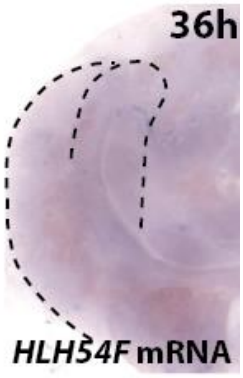

Figure 3.12. *HLH54F* is not expressed in the lateral plate during posterior lobe development.

in situ hybridization for *HLH54F* reveals no activity during genital development.

I next performed an *in situ* hybridization to observe expression of *Doc2*, one of the three *Doc* genes that comprise the *Doc* complex. *Doc2* expression occurs early in genital development at the border between the anal plate and genital arch, continuing along the border between the lateral plate and the clasper (Figure 3.13A, arrow). This "genital border pattern" is very similar to the genital pattern of Engrailed at this time in development (Glassford et al., 2015). Although the spatial extent of the *Doc2* pattern seems to be reduced in late *D. biarmipes* genitalia (Figure 3.13D), a more comprehensive comparison of *Doc2* expression is required to confirm this difference.

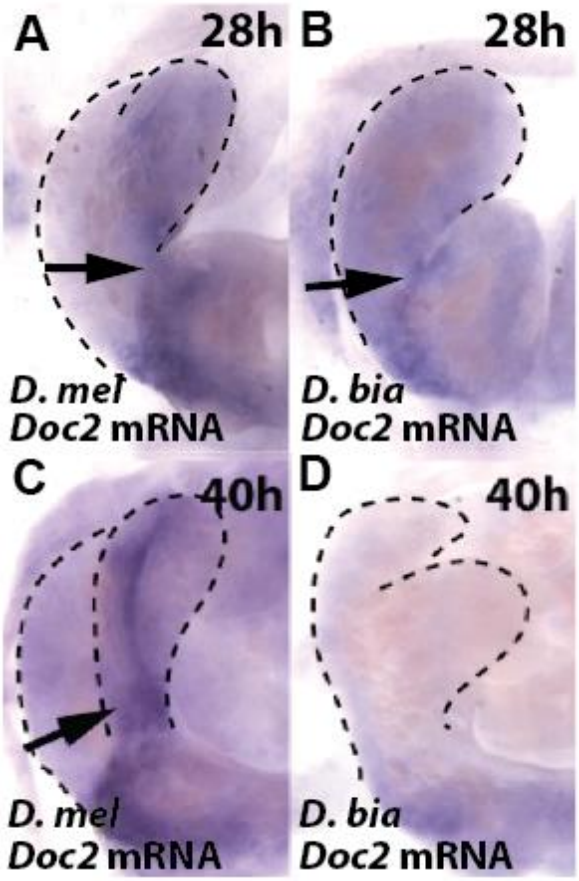

Figure 3.13. *Doc2* is expressed prior to and during the development of the posterior lobe.

(A-B) *in situ* hybridization for *Doc2* reveals *Doc2* expression near the posterior lobe tissue in both lobed (A, *D. melanogaster*) and non-lobed (B, *D. biarmipes*) species. (C-D) *Doc2* continues to be expressed in the clasper during later genital development in *D. melanogaster* (C), but not in *D. biarmipes* (D).

To assess the possibility that *Doc2* directly regulates *Delta,* I next screened the *Delta* posterior lobe enhancer region for *Doc2* consensus binding sites. I identified a predicted strong *Doc2* binding site that is conserved to *D. ananassae* (Figure 3.14A) within a region overlapped by three subfragments of the *Delta* enhancer (Figure 3.14B), all of which report the posterior lobe pattern (Figure 3.14D-F). None of these fragments of *Delta*'s posterior lobe enhancer drive expression in the LVM, suggesting that regulatory information necessary for LVM activity is

spread across the full ~2kb region (Figure 3.14D'-F'). Nevertheless, the posterior lobe and LVM activities of this region may be composed of overlapping, but partially distinct inputs.

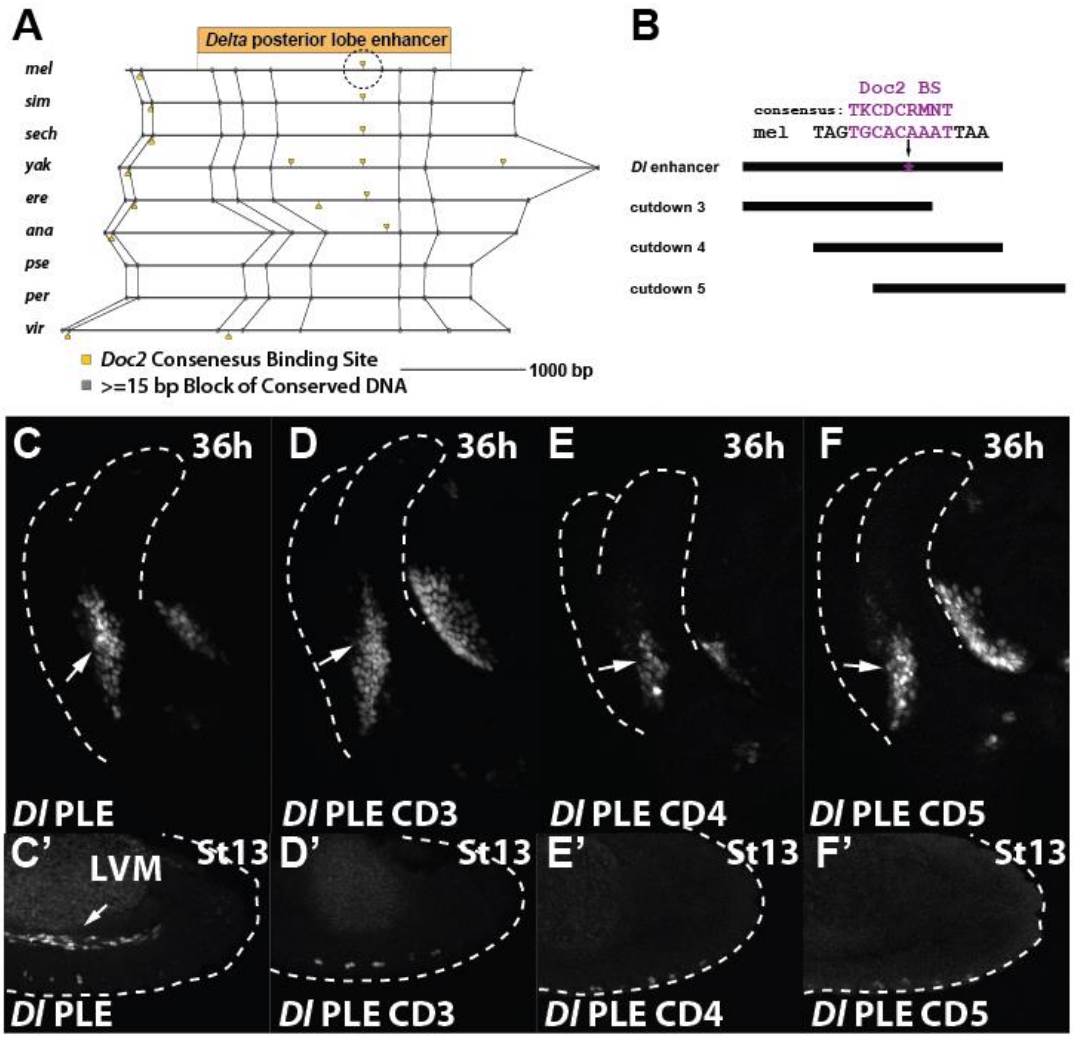

Figure 3.14. A region containing a predicted Doc2 binding site is necessary for the posterior lobe associated pattern of the *Delta* posterior lobe enhancer (PLE).

(A) Alignment of orthologous sequences from several lobed and non-lobed species covering the *Delta* PLE region (orange bar). (Yellow boxes) Predicted Doc2 binding sites (Gray boxes) Blocks of perfectly conserved sequence >=15 base pairs in length between listed *Drosophila* species (B) Schematic depicting location of predicted Doc2 binding site (purple asterisk) relative to three subfragments (cutdowns) of the *Delta* PLE. (C-F) Subfragments (cutdowns) of the *Delta* PLE that contain the predicted Doc2 binding site maintain the posterior lobe associated

pattern (D-F) compared to the full fragment (C). (C'-F') No cutdown maintains LVM activity (D'-F') compared to the full fragment (C').

Since *Doc2* is expressed in both the genitalia and LVM, I next screened the enhancers of *Doc2* to see if regulatory DNA is shared between the posterior lobe and LVM in a manner similar to the *Delta* posterior lobe enhancer. The *Doc* complex is ~40kb in length and contains three *Doc* paralogs: *Doc1, Doc2* and *Doc3* (Figure 3.15A). I tested each reporter associated with the *Doc* complex available from the Janelia Gal4 collection (Pfeiffer et al., 2008) for genital activity. Although a large portion of the regulatory regions of the *Doc* complex are covered by these reporters (Figure 3.15A, black bars), the published LVM enhancer is not included (Ismat et al., 2010). One reporter not associated with the LVM enhancer did exhibit a pattern that matched the *Doc2 in situ* hybridization pattern in the genitalia (Figure 3.15D-F, arrows). This enhancer also recapitulates endogenous *Doc2* expression patterns in several larval tissues including the leg, antennal and wing discs (not shown). Other reporters that contain fragments that overlap this *Doc2* "genital border enhancer" (GBE) do not report the early border pattern of *Doc2*, suggesting that the enhancer is restricted to a ~1kb region of DNA (Figure 3.15A). One overlapping reporter did recapitulate a late pattern of *Doc2* in the anal plate (Figure 3.15C and G, arrowheads), suggesting that this pattern is separable from the early genital border pattern. Although not complete, this analysis suggests that the genital and LVM activities of *Doc2* are separable and that its genital function likely arose independent from its LVM function.

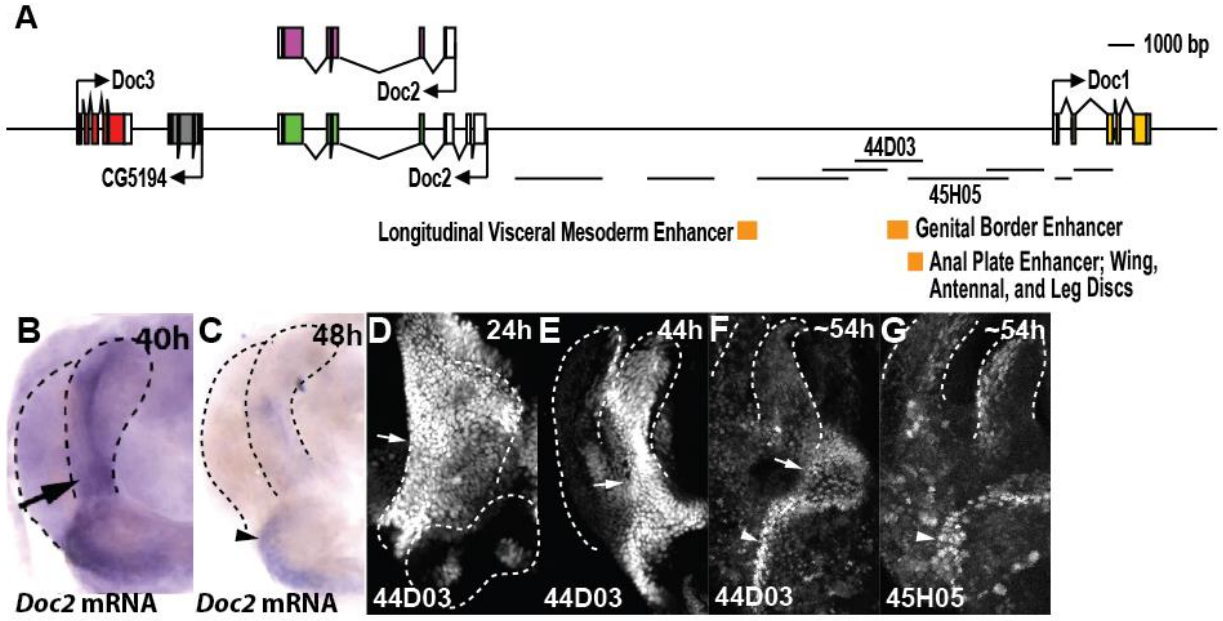

Figure 3.15. A screen of the regulatory regions of the *Doc* cluster.

(A) A schematic of the *Doc* cluster and its surrounding regulatory regions. Black lines refer to regions of regulatory DNA screened for genital activity in a transgenic reporter system. Orange lines depict known and observed activities of regulatory regions of *Doc*. (B-C) *in situ* hybridizations depicting *Doc2* expression in 40h (B) and 48h male pupal genitalia (D-F). Genital expression driven by a cross of Janelia Gal4 reporter line GMR44D03 (44D03) crossed to a UAS-nGFP reporter line at 24h (D), 44h (E) and 54h (F) (Pfeiffer et al., 2008). (G) Genital expression driven by a cross of Janelia Gal4 reporter line GMR45H05 (45H05) crossed to a UASnGFP fluorescent driver at 54h.

3.3 DISCUSSION

Here, we have identified Notch signaling as a key intercellular signaling pathway necessary for the development of the posterior lobe. This pathway is brought to the lobe primordium by a novel expansion of a pre-existing pattern of the Notch ligand *Delta* that is

unique to posterior lobed species (Figure 3.16A). Notch signaling is necessary for the upregulation of Jak/STAT signaling in the lobe, a key signaling event linked to the redeployment of several posterior spiracle network genes during posterior lobe development (Glassford et al., 2015). This suggests that a spatial shift in a pre-existing signaling pattern contributed to the co-option of the posterior spiracle network. Combined with my previous results which indicate the unlikely origins of much of the lobe network in the embryonic posterior spiracle, the findings presented here highlight how the assembly of a network can involve multiple parental networks that are connected through long chains of *trans*-regulatory connections. As few novelties have been resolved to this level of network mechanism, these results illuminate the current challenge of finding causative evolutionary events in vast networks.

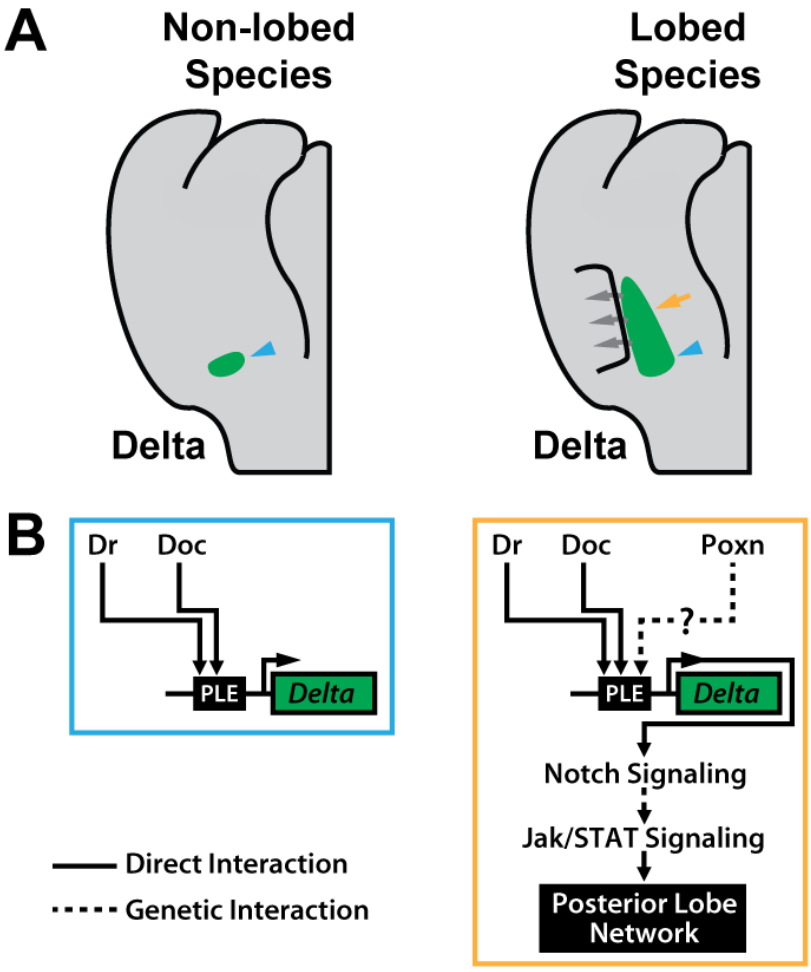

Figure 3.16. Model Depicting the role of Notch Signaling in the Evolution of the Posterior Lobe.

(A) Extent of lobe-associated *Delta* expression in non-lobed species (left) and lobed species (right). (blue arrowhead) Ancestral *Delta* pattern. (yellow arrow) novel expanded pattern of *Delta*. (B) Network diagrams illustrating the direct (black lines) and indirect (dotted lines) interactions within networks associated with the ancestral (left, blue outline) and novel (right, yellow outline) patterns of *Delta*. The novel expansion of *Delta* in lobed species requires *Poxn* activity, leads to the induction of the Notch and Jak/STAT signaling pathways and the subsequent deployment of the posterior lobe network (Glassford et al., 2015). The question mark indicates a possible interaction between *Poxn* and a hypothesized novel *trans* regulator of *Delta*.

Despite the discovery of several direct and indirect regulators of the *Delta* posterior lobe enhancer, the identity of this novel upstream factor remains unclear. *Drop* likely directly regulates the *Delta* posterior lobe enhancer thus integrating it with the genital sex-specification network (Chatterjee et al., 2011), but *Drop* is unlikely to be novel to lobed species. A predicted *Doc2* binding site in the *Delta* posterior lobe enhancer may also contribute to both ancestral and novel activities of *Delta*. The longer temporal persistence of *Doc2* mRNA transcript in *D. melanogaster* in comparison to the non-lobed *D. biarmipes* suggests that a difference in *Doc2* activity may contribute to the evolution of the posterior lobe, but a more extensive analysis is required. Cloning the orthologous regulatory regions of the *Doc* genital border enhancer into transgenic reporters would be key to resolving whether a *cis* modification of *Doc2* activity contributes to the expansion of *Delta*. Unlike *Drop* and *Doc*, *Poxn* is specifically necessary for the expansion of the lobe-associated *Delta* pattern (Figure 3.16B). Although we did not identify direct binding sites for *Poxn* in the *Delta* posterior lobe enhancer, this genetic interaction indicates that it is upstream of or collaborates with the novel factors(s) responsible for the expansion of *Delta* in lobed species. Regardless of the identity of the novel genes responsible for the origination of the posterior lobe network, we have identified a top-tier member of the posterior lobe network that gained a new activity through the reuse of pre-existing regulatory DNA. This finding mirrors the action of a previously-existing enhancer of *upd*, which increases the temporal extent of its expression in lobed species but contains no *cis*-regulatory alterations (Sarah Smith, personal communication). Together these results indicate that when a new expression pattern for a causative gene or genes evolve, it may ignite a chain of circuits with previous roles in unrelated networks.

3.4 METHODS

3.4.1 Fly Strains and Husbandry

All flies were reared on a standard cornmeal medium. Species used in this study were obtained from the UC San Diego *Drosophila* Stock Center (*Drosophila biarmipes* #0000-1028.01, *Drosophila ananassae* #0000-1005.01). The *Drosophila melanogaster* line used in this study is mutant for *yellow* and *white* (y^1w^1, Bloomington Stock Center #1495), and was isogenized for 8 generations.

3.4.2 Pupal Genital Sample Preparation

To collect developmentally staged genital samples, white prepupae were sorted by sex, and incubated at 25°C for 24 hours to 48 hours. Pupae were cut in half in cold PBS, extricated from the pupal case, and flushed with cold PBS to remove fat bodies and internal organs while preserving the developing genital epithelium. Carcasses were then fixed in PBS with 0.1% Triton-X and 4% paraformadehyde (PBT-fix) at room temperature for 30 minutes. Samples containing fluorescent reporters were washed three times for 10 minutes in PBS with 0.1% Triton-X (PBT) then imaged immediately. Samples to be used for *in situ* hybridization were rinsed twice in methanol and stored in ethanol at -20°C.

3.4.3 Embryo Collection

Embryos were collected from grape agar plates (Genesee Scientific) in egg-lay chambers that were incubated at 25°C for up to 20 hours. Embryos were dechorionated in 50% bleach for 3 minutes, washed in distilled water, and collected on a nitrile filter. Embryos were then fixed for 20 minutes in scintillation vials containing PBS, 2% paraformaldehyde, and 50% heptane. The PBS layer was removed from the vial and replaced with an equal amount of methanol. Samples to be used for *in situ* hybridization were vortexed for 30 seconds, removed from the methanol

layer, rinsed twice in methanol then stored in ethanol. Samples containing fluorescent reporters or to be used for immunostaining were shaken vigorously by hand for 1 minute, rinsed in methanol once then quickly rinsed in PBT three times to prevent the degradation of GFP and antibody epitopes.

3.4.4 Immunostaining

Embryo and genital samples were incubated overnight at 4°C with primary antibodies diluted in PBT. The following primary antibody were used: monoclonal mouse-anti-Dl (Developmental Studies Hybridoma Bank) and polyclonal goat-anti-Dl (dL-19) 1:100 Santa Cruz Biotechnology, Inc.) After several washes with PBT to remove unbound primary antibody, samples were incubated overnight in diluted secondary antibody (donkey anti-mouse Alexa 488 and donkey anti-goat Cy2), at 1:100 dilution from Molecular Probes to detect bound primary antibody. Samples were washed in PBT to remove unbound secondary antibody, incubated for 10 minutes in 50% PBT and 50% glycerol solution, then mounted on glass slides in an 80% glycerol 0.1M Tris-HCL 8.0 solution.

3.4.5 *in situ* Hybridization

in situ hybridization was performed as previously described in (Rebeiz et al., 2009) with the modification that we used an InsituPro VSi robot (Intavis Bioanalytical Instruments). Fixed embryo and genital samples were first dehydrated in a 50% xylenes/50% ethanol solution for 30 minutes at room temperature. Xylenes were removed by several washes with ethanol before the samples were loaded into the InsituPro VSi. During the automated steps, the samples were washed in methanol, rehydrated in PBT, fixed in PBT-fix, incubated in 1:25,000 proteinase K PBT (from a 10mg/mL stock solution), fixed in PBT-fix, and subjected to several washes in hybridization buffer. Samples were probed with a digoxygenin riboprobe targeting the coding

regions of selected genes (primers listed in Table 3.1) for 18 hours at 65°C. Unbound riboprobe was removed in several subsequent hybridization buffer washes, and washed several times in PBT. Samples were removed from the robot, and incubated overnight in PBT with 1:6000 anti-digoxygenin antibody Fab fragments conjugated to alkaline phosphatase (Roche Diagnostics). Alkaline phosphatase staining was then developed for several hours in NBT/BCIP color development substrate (Promega). Samples were then washed in PBT and mounted on glass slides in an 80% glycerol 0.1M Tris-HCL 8.0 solution.

3.4.6 Transgenic Constructs

Enhancer elements were cloned using the primers listed in the supplemental experimental procedure section, and inserted into the vector pS3aG (GFP reporter) using *AscI* and *SbfI* restriction sites as previously described (Williams et al., 2008). Primers were designed and sequence conservation was assessed using the GenePalette software tool (Rebeiz & Posakony, 2004). Targeted regions were cloned from genomic DNA purified using the DNeasy Blood & Tissue Kit (Qiagen). All GFP reporters were inserted into the 51D landing site on the 2nd chromosome (Bischof et al., 2007) by Rainbow Transgenics or the 68E1 landing site on the third chromosome (Bischof et al., 2007).

Delta-PLE→Delta-cDNA-pBAC vector was synthesized using In-Fusion cloning (Clontech) from PCR fragments cloned from *D. melanogaster* for the PLE and a *Delta* cDNA vector (Table 3.3) and combined with a pBAC vector backbone. The following GFP and Gal4 reporters were obtained from existing sources. Poxn-Gal4 (construct #13 from (Boll & Noll, 2002)) was obtained from Werner Boll, Transgenic RNAi lines from the Harvard TRiP project include *Delta*.

3.4.7 Microscopy

Adult posterior lobe cuticles and stained *in situ* hybridization samples were imaged on a Leica M205 stereomicroscope with a 1.6X objective with the extended multi-focus function. Samples stained with fluorescent antibodies or containing fluorescent reporters were imaged via confocal microscopy at 20X magnification on an Olympus Fluoview 1000 microscope. SEM images of third instar larvae were obtained as previously described by Higashijima (Higashijima et al., 1992).

For each transgenic construct, 3-5 independent lines inserted into the 51D landing site (Bischof et al., 2007) or 68A4 "attP2" landing site (Groth et al., 2004) were derived. We compared the relative expression of multiple lines in the genitalia to determine the normal reporter activity of each construct.

3.5 SUPPLEMENTAL TABLES

Gene	Species	Forward Primer	Reverse Primer
Drop	*D. melanogaster*	TACTTCCATCCGAGCCTGAT	taatacgactcactatagggagaGCAAACGAATGCCGTTCGTT
Doc2	*D. melanogaster*	CAGAACGATCGCATCACCAAGCTG	taatacgactcactataggAGCARRCTCTGCATATTCTGCTG
Doc2	*D. biarmipes*	CAGAACGATCGCATCACCAAGCTG	taatacgactcactataggAGCARRCTCTGCATATTCTGCTG
HLH54F	*D. melanogaster*	AARACCAAGYTGCCCAACATTCCG	taatacgactcactataggTTACCATGCCGTGTCCGTTGTTGTG
bin	*D. melanogaster*	CAACACCATACTCAGTGCGAACGAC	taatacgactcactataggAGTAGGCGTAGATCTCGGAGAGCG
fkh	*D. melanogaster*	ATGAGCTACGCCAGCATGGGATC	taatacgactcactataggCATGTCGTACATCTTGATGTCCGCC
zfh1	*D. melanogaster*	GATGAACAGTATTAAGCTGCCCG	taatacgactcactataggCACCGGAATGCTCGTATTTGTG

Table 3.1. Primers for amplifying mRNA probes.

Lowercase letters represent tag sequences for the T7 (Reverse primer) RNA polymerases used for probe synthesis.

Construct	Forward Primer	Reverse Primer
mel Delta PLE	TTCCGggcgcgccCATATACCATACGCAATGGCCAGG	TTGCCcctgcaggCACAACTGCGTGAGAAGTTGCC
bia Delta PLE	TTCCGggcgcgccGCKCGAATCACTCAAATGTCACC	TTGCCcctgcaggCRTTGCGTGTAATTAAAACRCMARC
ana Delta PLE	TTCCGggcgcgccGCKCGAATCACTCAAATGTCACC	TTGCCcctgcaggCRTTGCGTGTAATTAAAACRCMARC
Mel Dl CD3	TTCCGggcgcgccCATATACCATACGCAATGGCCAGG	TTGCCcctgcaggGCATTAAAGTGGGATTGGTATTGCAC
Mel Dl CD4	TTCCGggcgcgccGCAGTGTGTACCCGAATCACAGTAC	TTGCCcctgcaggCACAACTGCGTGAGAAGTTGCC
Mel Dl CD5	TTCCGggcgcgccGCTTAAGATAGTAAAGATCGTGG	TTGCCcctgcaggCRTTGCGTGTAATTAAAACRCMARC

Dl DS Tile1	TTCCGggcgcgccCGCACTTGGCCAGAATGATCACTG	TTGCCcctgcaggAACTAATGGCTGCAGCTTCGGTAG
Dl DS Tile2	TTCCGggcgcgccTCCGGGATGCAAGAAGTCGCTTAG	TTGCCcctgcaggGAATGGTCTTGGCCGTCTCATGATC
Dl DS Tile3	TTCCGggcgcgccGGATCTCGATCAATCACCCAGCC	TTGCCcctgcaggGACATGGCCTCGCATTACATCGC
Dl DS Tile4	TTCCGggcgcgccCATATACCATACGCAATGGCCAGG	TTGCCcctgcaggCACAACTGCGTGAGAAGTTGCC
Dl DS Tile5	TTCCGggcgcgccTGCTAGGATGGTCGTCTACTCTCG	TTGCCcctgcaggCCATCAAGTGCACCTTGGACCAAC
Dl DS Tile6	TTCCGggcgcgccTGACTGGTTCATGATGCTCTGGAG	TTGCCcctgcaggGCGAGAAGCTTCTGTACAGCAGAC
Dl DS Tile7	TTCCGggcgcgccTTGACATGCATCGACAGCTGGC	TTGCCcctgcaggAGTGGGAAGCACTCATGGAGCTC

Table 3.2. Primers used for transgenic constructs.

Lowercase letters represent restriction sites for *AscI* (Forward primers) and *Sbf I* (Reverse primers) used for cloning.

Construct	Forward Primer cDNA	Reverse Primer cDNA	Forward Primer Enhancer	Reverse Primer Enhancer
mel Dl PLE → mel Dl cDNA	ccgggcgaattcgccggcgcgccGGTTACACCAGAAAAACGGTTC	ACATTTCAGCCATGCCGCATACAGGT	ACCTGTATGCGGCATGGCTGAAATGT	ttatgatctagagtcgcggccgcGTTATGGTCGTGTCAGACATCC

Table 3.3. Primers for generating infusion PCR fragments.

Lowercase letters indicate overlap with target vector. *Delta* pBAC construct was generated from 2 overlapping fragments combined with digest vector backbone.

Construct	Forward Primer	Reverse Primer
Delta Box 2 Mutant	GCAAAAAACTATGCGcGaCgAgTcAgTcGcTaAGCGGGCTTATG	CATAAGCCCGCTtAgCgAcTgAcTcGtCgCGCATAGTTTTTTGC

Table 3.4. Primers for generating *Delta* box 2 mutation.

Lowercase letters indicate altered sites introduced using these primers. *Delta* pBAC construct was generated from 2 overlapping fragments combined with digest vector backbone.

4.0 THE MODIFICATION OF A CO-OPTED ENHANCER CONTRIBUTES TO THE DIVERSIFICATION OF TWO EXTERNAL GENITAL STRUCTURES

The following collaborators contributed data presented in this chapter: Winslow Johnson and Mark Rebeiz cloned the *Poxn* PLE reporters (Figure 4.3B-G). Mark Rebeiz cloned the *Poxn* transgenes (Figure 4.1K-N). Rachel Pileggi collected the SEM images (Figure 4.1A-B). Natalie Dall assisted with cloning the *Poxn* PLE box mutants, and collected and analyzed all images of the *Poxn* PLE box mutant reporters (Figure 4.A1).

4.1 INTRODUCTION

In previous chapters, it was revealed that several posterior lobe network genes are deployed in the genitalia by enhancers that also have activities in other tissues during development. The co-option and redeployment of regulatory DNA provides a rapid means for the construction of novel developmental networks, addressing the issue of how complex novelties may be quickly assembled while avoiding multiple non-functional intermediate steps (Monteiro & Podlaha, 2009). One major conundrum concerning the network co-option mechanism however is that it is predicted to create pleiotropic constraint between the derived and ancestral networks (Wagner & Zhang, 2011). **Pleiotropy** occurs when one gene or genetic variant influences two or more phenotypic traits (Stearns, 2010; Wagner & Zhang, 2011). Because a co-opted enhancer is active

in two or more tissues, its modification may cause pleiotropic effects in other tissues. Currently, little is known about the molecular mechanisms that allow co-opted networks to diverge from their ancestral functions.

The process of network individualization is critical to understand because novel morphological structures frequently diversify into a wide variety of forms after their inception (Erwin, 2015). The butterfly wingspot is a bullseye-shaped pattern that greatly varies in color, position and number of rings throughout the nyphalid family, but is thought to have originated from a common ancestor (Oliver et al., 2012). Thousands of beetle species exhibit horns that vary in number, position, shape and size (Moczek & Nagy, 2005), and is another morphological novelty that likely originated from a common ancestor. Evidence suggests that network co-option contributed to the evolution to both the beetle horn and butterfly eyespot (Keys et al., 1999; Moczek & Nagy, 2005), but it is mechanistically unclear how these novelties were able to diversify under the proposed pleiotropic constraint of network co-option.

One mechanism proposed to contribute to the individualization of co-opted networks is the evolution of separate *cis*-regulatory elements unique to each developmental context (Monteiro & Podlaha, 2009). Regulatory DNA may be especially amenable to rapid separation due to the persistent turnover of enhancer components (Swanson et al., 2011). An alternative hypothesis would be the modulation of upstream tiers in a novel network's hierarchy (Erwin & Davidson, 2009), as certain types of genes in networks have been proposed to be the preferential targets of evolution as they uniquely control co-operating suites of genes (Stern & Orgogozo, 2008). To understand how co-opted networks can be diversified, examples of recently co-opted networks that have been subsequently diversified are needed. Further, these must be worked out

in systems for which the regulatory elements can be identified ,and connected to their phenotypic outcomes evolutionarily.

The posterior lobe is a recently derived novelty in the *Drosophila melanogaster* clade (Kopp & True, 2002). This structure has been shown to contact the female ovipositor during copulation, and is necessary for successful copulation (Frazee & Masly, 2015; Kamimura, 2010). Like many external sexual structures (Eberhard, 1985), the posterior lobe has rapidly diverged in shape and size, including *D. mauritania*, *D. sechellia*, and *D. simulans*, which last shared a common ancestor 300,000-900,000 years ago (Tamura et al., 2004). Interbreeding these species yields fertile F1 females, allowing quantitative trait loci analysis, which has revealed that numerous loci underlie the differences between these species (J. A. Coyne, Rux, & David, 1991; Laurie, True, Liu, & Mercer, 1997; Macdonald & Goldstein, 1999; Zeng et al., 2000). We recently discovered 7 posterior lobe genes that were co-opted into the posterior lobe network through the reuse of pre-existing enhancers (Glassford et al., 2015). As it is possible that the co-opted genes of the posterior spiracle were subsequently modified to contribute to the diversification of the posterior lobe, this represents an excellent system in which to disentangle how co-opted networks are individualized.

Here, we study the contribution of one co-opted posterior spiracle network gene, *Poxn*, to the evolution of two external male genital structures: the posterior lobe and the clasper. We demonstrate that *Poxn* transgenes cloned from *D. simulans* and *D. mauritiana* differentially rescue *Poxn* mutant morphologies of the clasper and posterior lobe. Transgenic reporters of the posterior lobe enhancer (PLE) region of *Poxn* reveal that a *cis*-regulatory change contributes to both the clasper and posterior lobe phenotypes. Analysis of the shared posterior spiracle activity of the *Poxn* PLE reveals that this modification does not disrupt the embryonic pattern. These

results demonstrate that despite the reuse of shared regulatory information, co-opted enhancers may be modified while avoiding a pleiotropic constraint.

4.2 RESULTS

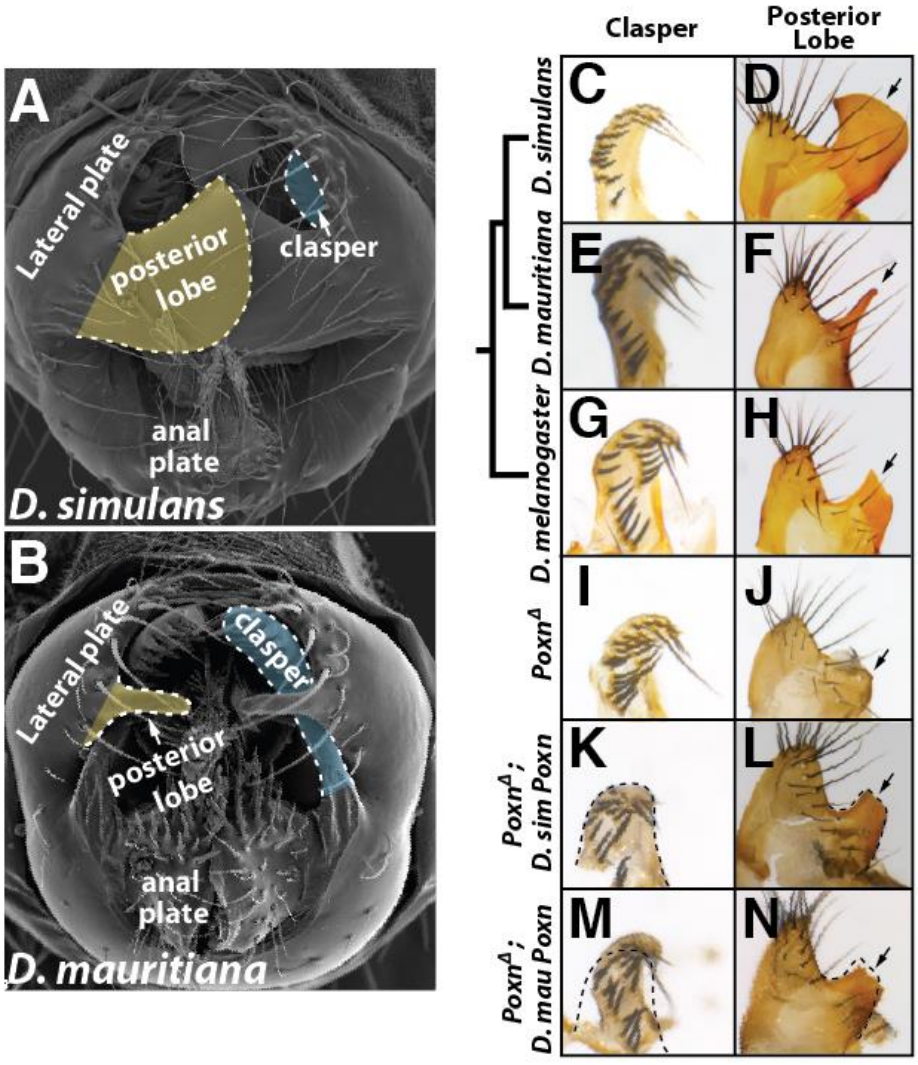

Figure 4.1. *Poxn* contributes to the evolution of two external male genital appendages.

(A-B) Scanning electron micrographs of *D. simulans* (A) and *D. mauritiana* (B) reveal divergent male genital morphologies. Lateral plate, anal plate, clasper (highlighted in blue), and posterior lobe (highlighted in

yellow) structures are labeled. (C-H) Images of dissected claspers (C,E,G) and lateral plates (D,F,H) isolated from the selected species within the *D. melanogaster* clade. (I-N) Clasper (I) and posterior lobe (N) from *D. melanogaster* flies containing the a null $Poxn^{\Delta M22\text{-}B5^{=}}$allele (Boll & Noll, 2002) are altered in shape and reduced in size. (K-N) *Poxn* transgenes cloned from *D. simulans* and *D. mauritiana* differentially rescue the clasper (K, M) and posterior lobe (L, N) phenotypes. Dotted black lines outline the *D. simulans Poxn* rescue phenotype.

4.2.1 Differences at *Poxn* Contribute to the Evolution of the Posterior Lobe

In addition to a complex intromittent organ, the male genitalia of the *Drosophila melanogaster* clade species bear several secondary external cuticular structures including the posterior lobe, lateral plates, anal plates, and a bristle-laden grasping appendage named the clasper (Figure 4.1A-B). The posterior lobe and clasper exhibit the most drastic morphological differences between the sister species *D. simulans* and *D. mauritiana. D. mauritiana's* posterior lobe exhibits an altered shape and is greatly reduced in size compared to *D. simulans* (Figure 4.1D and F). In contrast, the clasper of *D. mauritiana* is larger than that of *D. simulans* and exhibits altered bristle morphology (Figure 4.1C and E). Flies containing a null mutation for the paired domain transcription factor *Pox-neuro (Poxn)* exhibit defective posterior lobe and clasper morphologies (Figure 4.1I-J) compared to a wild-type *D. melanogaster* (Boll & Noll, 2002). These defects can be rescued by *Poxn* transgenes containing an intronic genital enhancer of the gene (Glassford et al., 2015). Given the fundamental requirement of *Poxn* for the construction of the posterior lobe, we explored the possibility that variation at this key gene may have contributed to its diversification.

To test whether alterations at *Poxn* contributed to the evolution of differences in the posterior lobe and clasper, we cloned a 9 kilobase genomic fragment of *Poxn*, spanning its

coding and intronic DNA from *D. simulans* and *D. mauritiana*, and compared their ability to transgenically complement the genital defects of the null $Poxn^{\Delta M22\text{-}B5}$ mutant. Both transgenes rescued the posterior lobe and clasper *Poxn* mutant phenotypes, but to differing degrees (Figure 4.1K-N). The claspers of *Poxn* mutant flies containing the *D. simulans Poxn* transgene were smaller than those containing the *D. mauritania* transgene. The posterior lobes of *Poxn* mutant flies containing the *D. simulans Poxn* transgene were approximately 11.0% larger than those containing the *D. mauritiana* transgene. These results implicate modifications of *Poxn* during the diversification of the lobe and clasper structures.

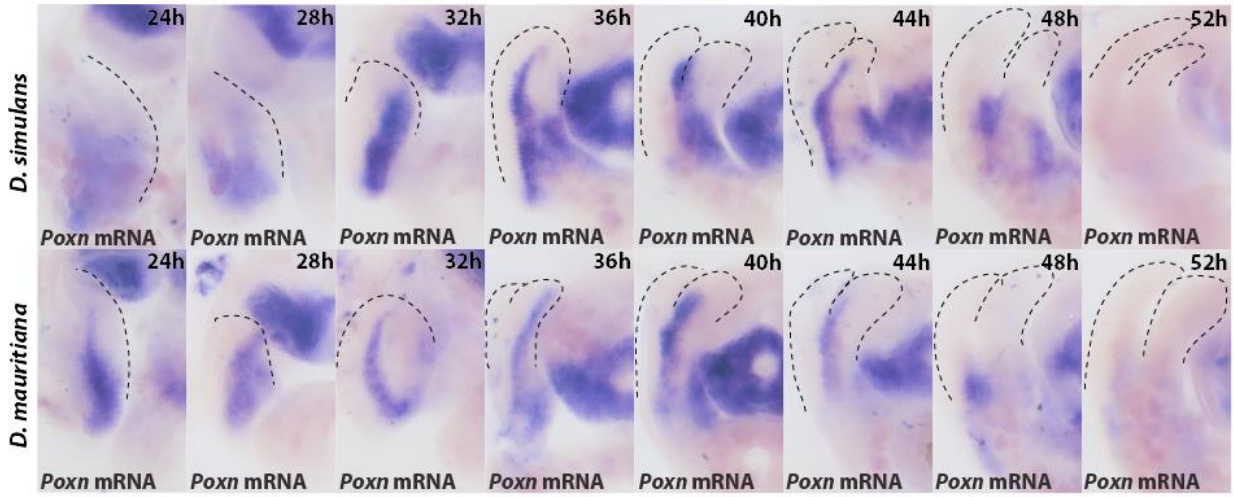

Figure 4.2. *Poxn* is differentially expressed between *D. simulans* and *D. mauritiana* during genital development.

in situ hybridization for *Poxn* mRNA in pupal male genitalia reveal higher levels of transcript during the cleavage of the clasper from the lateral plate (32h) and at the base of the developing posterior lobe (36h to 44h). Dotted black lines outline the lateral plate and clasper tissues.

4.2.2 *Cis*-Regulatory Changes at *Poxn* Contribute to Posterior Lobe and Clasper Evolution

The difference in transgenic complementation of the *D. simulans* and *D. mauritiana Poxn* constructs suggested that this region contains mutations that have contributed to phenotypic changes in these species' genitalia. This could be due to mutations in the cis-regulatory sequences contained within the construct, or a change in the coding region of *Poxn.* Sequencing of the rescue constructs revealed a single leucine/isoleucine difference between *D. simulans* and *D. mauritiana* in the *Poxn* protein-coding region. As this represents a conservative change predicted to minimally alter the protein's structure, we next sought to ascertain whether the differential rescue of the *Poxn* mutant phenotype may have been caused by a change in gene expression. *Poxn* is expressed in two consecutive waves in the lateral plate epithelium during pupal development, the first wave is expressed during the cleavage of the clasper from the lateral plate, and a second wave, unique to species with posterior lobes, is expressed at the base of the developing posterior lobe (Glassford et al., 2015). We observed higher levels of *Poxn* mRNA transcript in *D. simulans* compared to *D. mauritiana* during the cleavage of the clasper from the lateral plate (32h) and during the emergence of the posterior lobe from the lateral plate (36h-44h) in an *in situ* hybridization experiment (Figure 4.2). Antibody staining for Poxn protein revealed that this difference is correlated with a difference in protein level: Poxn protein accumulates to higher levels in *D. simulans* compared to *D. mauritiana* in both the clasper associated pattern (Figure 4.3A and C) and the posterior lobe associated pattern (Figure 4.3B and D). This suggests that the differences in transgenic complementation may be caused by changes in *Poxn* expression.

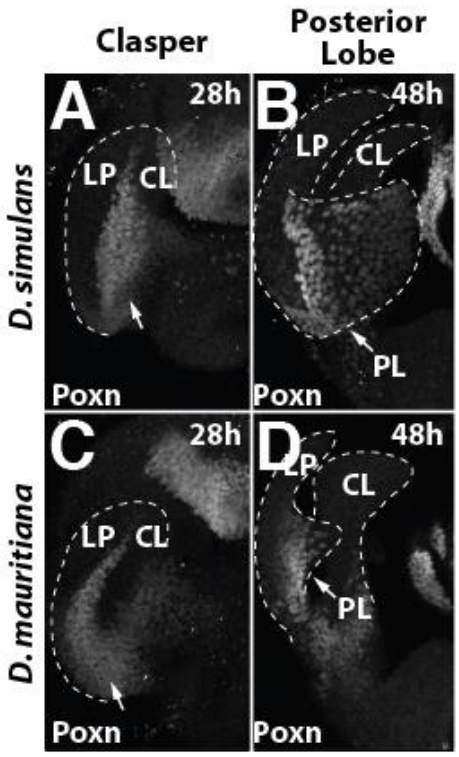

Figure 4.3. *Poxn* exhibits differential activity in the male genitalia in patterns associated with the clasper and posterior lobe.

Antibody staining for Poxn reveals higher Poxn activity in early genitalia in *D. simulans* compared to *D. mauritiana* (A and C, arrows), and higher *Poxn* activity at the base of the developing posterior lobe in later genital development in *D. simulans* compared to *D. mauritiana* (B and D, arrows). LP: lateral plate, CL: clasper, PL: posterior lobe.

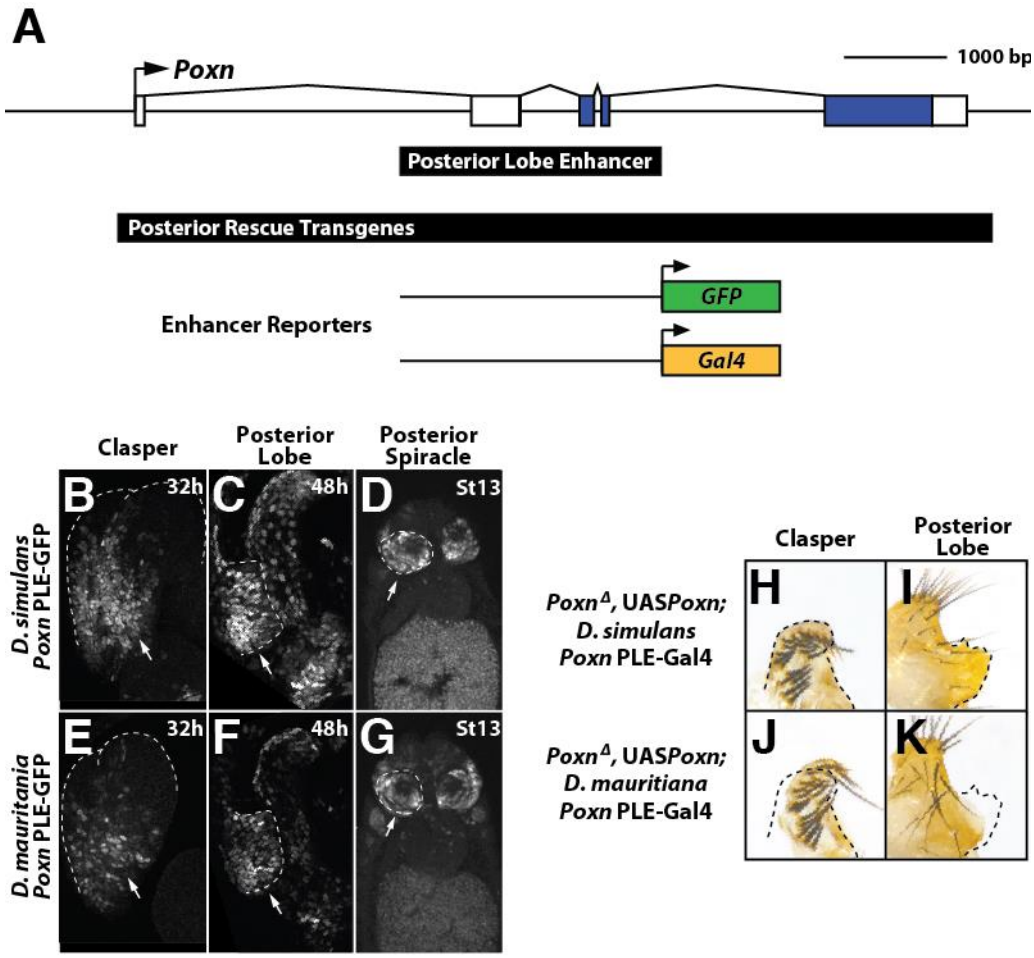

Figure 4.4. The *Poxn* posterior lobe enhancer (PLE) contributes to clasper and posterior lobe evolution without disrupting its posterior spiracle activity.

(A) Schematic of the *Poxn* gene locus, showing the relative position of the *Poxn* PLE region displaying the relative positions of rescue fragments, green fluorescent protein (GFP), and *Gal4* transgenic constructs. (B-G) Transgenic GFP reporters of the orthologous *Poxn* PLE from *D. simulans* and *D. mauritiana* reveal differential activity in patterns associated with the clasper (B and E) and the posterior lobe (C and F), but not in the posterior spiracle (D and G). (H-K) Phenotypic rescue of a *Poxn* mutant phenotype by *D. simulans* and *D. mauritiana Poxn-PLE Gal4* constructs driving a UAS-*Poxn* transgene results in contrasting clasper (H and J) and posterior lobe (I and K) morphologies. Dotted black lines outline the *D. simulans Poxn* PLE rescue phenotype.

The observed differences in *Poxn* expression, and lack of major alterations in its coding region indicated that the phenotypically relevant variation at *Poxn* may be caused by *cis* changes to its transcriptional regulatory sequences (Figure 4.1K-N). Previously, I characterized a ~2.6kb enhancer region of *Poxn* that spans its second exon and intron (Figure 4.4A) (Glassford et al., 2015). This posterior lobe enhancer (PLE) recapitulates endogenous *Poxn* expression during both waves of its deployment in the presumptive clasper/lateral plate border and posterior lobe (Glassford et al., 2015). To determine whether the differences in the *D. simulans* and *D. mauritiana Poxn* transgenes reside within this enhancer region, we cloned the orthologous *Poxn* PLEs from the two species into green fluorescent protein (GFP) reporter constructs. Animals bearing the *D. simulans Poxn* PLE reporter exhibited higher levels of expression during the cleavage of the clasper from the lateral plate and during posterior lobe development (Figure 4.4B and C) compared to those bearing a *D. mauritiana Poxn* PLE reporter (Figure 4.4E and F). This data indicates that the posterior lobe enhancer of *Poxn* was altered following its co-option to the lobe.

While the differences in reporter activity correlate with the contrasting phenotypes of *D. simulans* and *D. mauritiana* rescue transgenes, we sought further evidence that these differences could be explained by the transcriptional activity of this region, as the PLE overlaps both coding and non-coding exonic sequences (Glassford et al., 2015). To isolate changes in transcriptional mechanisms from alterations to post-transcriptional or coding sequences, we performed a transgenic complementation assay in which only transcriptional mechanisms could contribute phenotypic differences. We cloned *D. simulans* and *D. mauritiana Poxn* PLEs upstream of the GAL4 transcription factor and inserted these constructs into a common landing site on the third chromosome. These driver lines were crossed into a background homozygous for the null *Poxn*

mutant which contained a UAS-*Poxn* transgene. In this experiment, any expression differences in *Poxn* would be due to the transcriptional activity of the *Poxn* PLE inserted upstream of the GAL4 gene. Both the clasper and posterior lobe phenotypes were differentially rescued by the two reporters: flies containing the *D. simulans Poxn* PLE exhibited claspers larger than flies with the *D. mauritiana*'s *Poxn* PLE (Figure 4.4H and J), and the posterior lobes of flies containing the *D. simulans Poxn* PLE were larger than those from flies containing the *D. mauritiana Poxn* PLE *Gal4* construct (Figure 4.4I and K). These data indicate that the observed differences in the *Poxn* PLE activity are phenotypically relevant, and consistent with the differences observed in the transgenic complementation assay using genomic rescue constructs (Fig 4.1K-N).

4.2.3 The Diversification of the *Poxn* PLE Preserved an Ancestral Function

Given the functional differences localized to the PLE in the experiments above, a key question is whether the mutations in this element had effects that extended beyond the genital functions of this region. Previously, I showed how this enhancer also drives expression in, the embryonic posterior spiracle as a result of the co-option of posterior spiracle network to lobe development (Glassford et al., 2015). Considering that all three activities promoted by the *Poxn* PLE utilize at least two of the same transcription factor binding sites (Glassford et al., 2015), I sought to determine whether the modification of the posterior lobe and clasper activities incurred a pleiotropic effect on the embryonic expression pattern. Comparing stage 13 embryos bearing the *D. simulans* or *D. mauritiana* PLE reporters, we did not observe a significant difference in posterior spiracle-associated activity (Figure 4.4D and G). This indicates that modification to the clasper and posterior lobe-associated activities of the *Poxn* PLE did not disrupt its posterior spiracle activity.

4.3 DISCUSSION

Here, we have identified a modified enhancer that contributes to the diversification of two external male genital morphologies, the posterior lobe and clasper. Although differences in the posterior lobe have been appreciated for many years (Bock & Wheeler, 1972; Kopp & True, 2002), and subject to quantitative trait loci analysis (J. Coyne, 1993; Tanaka et al., 2015) our results represent the first documented case in which causative changes have been localized to a specific gene or mechanism. While this modification altered the level of transcription driven by this enhancer in two genital patterns, it did not disrupt a third activity in the embryonic posterior spiracle, which nevertheless shares key inputs with the clasper and posterior lobe. These results illustrate how pleiotropic constraint on co-opted circuits may be circumvented, and yet raise a possible limitation to how modular such circuits may be. The posterior lobe-associated activity is deployed in species with posterior lobes through the reuse of pre-existing regulatory DNA (Glassford et al., 2015), which may have incurred a pleiotropic constraint between it and the clasper-associated activity. This pleiotropic constraint did not extend to the posterior spiracle pattern, perhaps due to a higher degree of similarity in the *trans*-regulatory landscape between genital tissues compared to the embryo. I briefly discuss the possible molecular mechanisms that may mediate this phenomenon and its potential implications below.

Although several posterior spiracle network genes were co-opted into the posterior lobe (Glassford et al., 2015), many differences in the *trans* regulatory landscape likely persist between the embryonic and genital tissues. One mechanism to introduce modularity into a shared enhancer is the addition of a binding site for a tissue-specific regulator. In a hypothetical model for the evolution of *Poxn*, a repressor is uniquely expressed in a portion of the genitalia that includes the posterior lobe (Figure 4.5A). After the divergence separating *D. simulans* from *D.*

mauritiana, a mutation could introduce a binding site for the repressor into one of the sister species' posterior lobe enhancers, specifically reducing expression of *Poxn* in the posterior lobe (Figure 4.5B). The potential for any enhancer to gain binding sites without disrupting any of its ancestral activities likely depends on the nature and composition of the enhancer. Enhancers may be spread over thousands of bases and exhibit high DNA turnover (Swanson et al., 2011), or conserved, leaving little opportunity for modification.

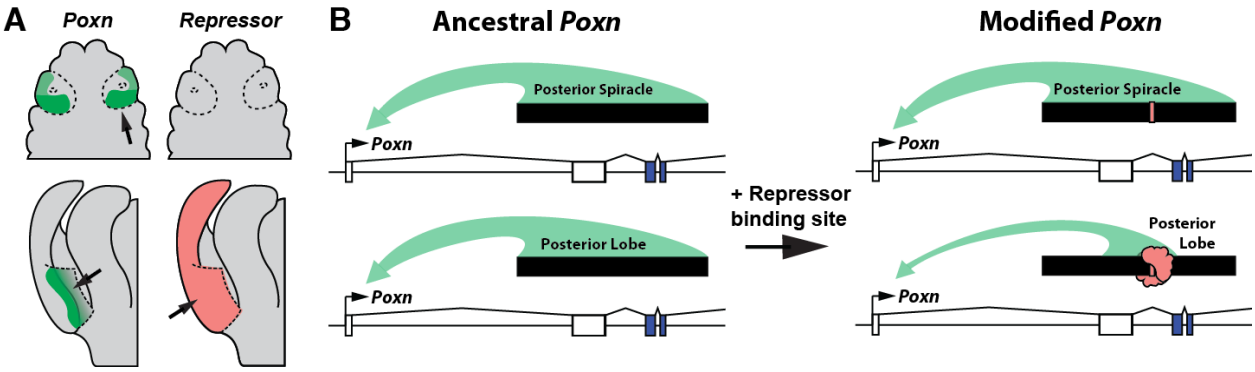

Figure 4.5. Hypothetical Model for the Individualization of *Poxn*.

(A) *Poxn* is expressed in the posterior spiracle (top, green), and posterior lobe (bottom, green). A hypothetical repressor is expressed in a portion of the genitalia that includes the posterior lobe (bottom, red), but not in the posterior spiracle. (B) Schematic of *Poxn* gene locus and *Poxn* posterior lobe enhancer region (black bar). The introduction of a binding site for the genital-specific repressor (red box) specifically reduces expression induced by the *Poxn* posterior lobe enhancer (green) in the posterior lobe due to the lobe-specific activity of the repressor.

Several hypotheses have been presented as the driving force behind the rapid evolution of external male genitalia (Eberhard, 1985; Reinhardt, 2009), including cryptic female selection, sexually antagonistic coevolution, natural selection, lock-and-key coevolution and pleiotropy. Studying the molecular mechanisms causing pleiotropy-driven morphological evolution can be difficult, as it requires knowledge of the underlying genetic links between tissues. Morphologies

generated by network co-option present the opportunity to investigate such genetic links between the novel and ancestral structures. In the case of the difference in genital activity between *D. simulans* and *D. mauritiana* at *Poxn*, while there seemed to be no pleiotropic alteration of the posterior spiracle, both the clasper and posterior lobe are shaped by a change in a shared enhancer. As both structures are part of the external male genitalia this finding raises the possibility that pressure to evolve the clasper or lobe activity may have caused the other to passively change due to pleiotropic constraint. A QTL analysis of several external male genital structures found little connection between genes that modified the posterior lobe and the clasper (Tanaka et al., 2015), suggesting that these structures are not generally pleiotropically linked at other contributing loci. Thus, further study of co-opted genes may shed light on the role of pleiotropy in driving the rapid evolution of individual male external genital structures. The *Poxn* posterior lobe enhancer represents the first characterized gene that contributes to the evolution of external male genital morphology in insects, and represents a valuable tool for studying selective forces guiding the rapid evolution of the external male genitalia.

4.4 METHODS

4.4.1 Fly Strains and Husbandry

All flies were reared on a standard cornmeal medium. Species used in this study were obtained from the UC San Diego *Drosophila* Stock Center (*Drosophila simulans* #14021-0251.165, *Drosophila mauritiana* #14021-0241.01). The *Drosophila melanogaster* line used in

this study is mutant for *yellow* and *white* (y^1w^1, Bloomington Stock Center #1495), and was isogenized for 8 generations.

4.4.2 Pupal Genital Sample Preparation

To collect developmentally staged genital samples, white prepupae were sorted by sex, and incubated at 25°C for 24 hours to 48 hours. Pupae were cut in half in cold PBS, extricated from the pupal case, and flushed with cold PBS to remove fat bodies and internal organs while preserving the developing genital epithelium. Carcasses were then fixed in PBS with 0.1% Triton-X and 4% paraformadehyde (PBT-fix) at room temperature for 30 minutes. Samples containing fluorescent reporters were washed three times for 10 minutes in PBS with 0.1% Triton-X (PBT) then imaged immediately. Samples to be used for *in situ* hybridization were rinsed twice in methanol and stored in ethanol at -20°C.

4.4.3 Embryo Collection

Embryos were collected from grape agar plates (Genesee Scientific) in egg-lay chambers that were incubated at 25°C for up to 20 hours. Embryos were dechorionated in 50% bleach for 3 minutes, washed in distilled water, and collected on a nitrile filter. Embryos were then fixed for 20 minutes in scintillation vials containing PBS, 2% paraformaldehyde, and 50% heptane. The PBS layer was removed from the vial and replaced with an equal amount of methanol. Samples to be used for *in situ* hybridization were vortexed for 30 seconds, removed from the methanol layer, rinsed twice in methanol then stored in ethanol. Samples containing fluorescent reporters or to be used for immunostaining were shaken vigorously by hand for 1 minute, rinsed in

methanol once then quickly rinsed in PBT three times to prevent the degradation of GFP and antibody epitopes.

4.4.4 Immunostaining

Embryo and genital samples were incubated overnight at 4°C with primary antibodies diluted in PBT. The following primary antibody were used: rabbit-anti-Poxn 1:100 (Dambly-Chaudiere et al., 1992) After several washes with PBT to remove unbound primary antibody, samples were incubated overnight in diluted secondary antibody (donkey anti-rabbit Alexa 488), both at 1:400 dilution from Molecular Probes to detect bound primary antibody. Samples were washed in PBT to remove unbound secondary antibody, incubated for 10 minutes in 50% PBT and 50% glycerol solution, then mounted on glass slides in an 80% glycerol 0.1M Tris-HCL 8.0 solution.

4.4.5 *in situ* Hybridization

in situ hybridization was performed as previously described in (Rebeiz et al., 2009) with the modification that we used an InsituPro VSi robot (Intavis Bioanalytical Instruments). Fixed embryo and genital samples were first dehydrated in a 50% xylenes/50% ethanol solution for 30 minutes at room temperature. Xylenes were removed by several washes with ethanol before the samples were loaded into the InsituPro VSi. During the automated steps, the samples were washed in methanol, rehydrated in PBT, fixed in PBT-fix, incubated in 1:25,000 proteinase K PBT (from a 10mg/mL stock solution), fixed in PBT-fix, and subjected to several washes in hybridization buffer. Samples were probed with a digoxygenin riboprobe targeting the coding

regions of selected genes (primers listed in Table 4.1) for 18 hours at 65°C. Unbound riboprobe was removed in several subsequent hybridization buffer washes, and washed several times in PBT. Samples were removed from the robot, and incubated overnight in PBT with 1:6000 anti-digoxygenin antibody Fab fragments conjugated to alkaline phosphatase (Roche Diagnostics). Alkaline phosphatase staining was then developed for several hours in NBT/BCIP color development substrate (Promega). Samples were then washed in PBT and mounted on glass slides in an 80% glycerol 0.1M Tris-HCL 8.0 solution.

4.4.6 Transgenic Constructs

Enhancer elements were cloned using the primers listed in the supplemental experimental procedure section, and inserted into the vector pS3aG (GFP reporter) or pS3aG4 (Gal4 reporter) using *AscI* and *SbfI* restriction sites as previously described (Williams et al., 2008). Primers were designed and sequence conservation was assessed using the GenePalette software tool (Rebeiz & Posakony, 2004). Targeted regions were cloned from genomic DNA purified using the DNeasy Blood & Tissue Kit (Qiagen). All GFP reporters were inserted into the 51D landing site on the 2nd chromosome (Bischof et al., 2007) by Rainbow Transgenics. Gal4 insertions depicted in Figure S3 were inserted into the 68E1 landing site on the third chromosome (Bischof et al., 2007). A full list of transgenes and insertions sites is listed in Table 4.4.

The *Poxn* rescue construct were synthesized using In-Fusion cloning (Clontech) from 2 ~4.5 PCR fragments covering the coding and intronic regions of *Poxn* from *D. simulans* and *D. mauritiana* (Table 4.3) and pS3aG vector backbone digested with *AscI* and *NotI* restriction enzymes. The following GFP and Gal4 reporters were obtained from existing sources. Poxn-Gal4 (construct #13 from (Boll & Noll, 2002)) and UAS-Poxn was obtained from Werner Boll.

4.4.7 Microscopy

Adult posterior lobe cuticles and stained *in situ* hybridization samples were imaged on a Leica M205 stereomicroscope with a 1.6X objective with the extended multi-focus function. Samples stained with fluorescent antibodies or containing fluorescent reporters were imaged via confocal microscopy at 20X magnification on an Olympus Fluoview 1000 microscope. SEM images of third instar larvae were obtained as previously described by Higashijima (Higashijima et al., 1992).

For each transgenic construct, 3-5 independent lines inserted into the 51D landing site (Bischof et al., 2007) or 68A4 "attP2" landing site (Groth et al., 2004) were derived. A list of reporters and corresponding landing sites are reported in Table 4.4. We compared the relative expression of multiple lines in the genitalia to determine the normal reporter activity of each construct. For quantitative measures, relative fluorescence of the were determined in both the posterior lobe and posterior spiracle contexts. Mounted genital and embryo samples were imaged at 40X magnification under identical, non-saturating settings uniquely optimized for each sample type. Relative expression within the lobe or spiracle was quantified using ImageJ and assessed using a student's paired t-test.

4.5 SUPPLEMENTAL TABLES

Gene	Species	Forward Primer	Reverse Primer
poxn	*D. simulans*	ACCGTGGTGAAGAAGGATCATCC	taatacgactcactataggCAGATCAAAACTGGGTCAGTGG
poxn	*D. mauritiana*	ACCGTGGTGAAGAAGGATCATCC	taatacgactcactataggCAGATCAAAACTGGGTCAGTGG

Table 4.1. Primers for amplifying mRNA probes.

Lowercase letters represent tag sequences for the T7 (Reverse primer) RNA polymerases used for probe synthesis.

Construct	Forward Primer	Reverse Primer
sim poxn PLE	TTCCGggcgcgccTCGGTGGCTTAACACGCGCATT	TTGCCcctgcaggATCGCTGATTCCATGGCCCAGT
mau poxn PLE	TTCCGggcgcgccTCGGTGGCTTAACACGCGCATT	TTGCCcctgcaggATCGCTGATTCCATGGCCCAGT
sim poxn PLE Gal4	TTCCGggcgcgccTCGGTGGCTTAACACGCGCATT	TTGCCcctgcaggATCGCTGATTCCATGGCCCAGT
mau poxn PLE Gal4	TTCCGggcgcgccTCGGTGGCTTAACACGCGCATT	TTGCCcctgcaggATCGCTGATTCCATGGCCCAGT

Table 4.2. Primers used for transgenic constructs.

Lowercase letters represent restriction sites for *AscI* (Forward primers) and *Sbf I* (Reverse primers) used for cloning

Construct	Forward Primer A	Reverse Primer A	Forward Primer B	Reverse Primer B
sim poxn rescue	ccgggcgaattcgccggcgcgccGGTTACACCAGAAAAACGGTTC	ACATTTCAGCCATGCCGCATACAGGT	ACCTGTATGCGGCATGGCTGAAATGT	ttatgatctagagtcgcggccgcGTTATGGTCGTGTCAGACATCC
mau poxn rescue	ccgggcgaattcgccggcgcgccGGTTACACCAGAAAAACGGTTC	ACATTTCAGCCATGCCGCATACAGGT	ACCTGTATGCGGCATGGCTGAAATGT	ttatgatctagagtcgcggccgcGTTATGGTCGTGTCAGACATCC

Table 4.3. Primers for generating infusion PCR fragments.

Lowercase letters indicate overlap with target vector. *Poxn* rescue constructs were generated from 2 overlapping fragments of the *Poxn* coding and intronic DNA.

Reporter Name	Species	51D9	68E
sim Poxn posterior lobe enhancer	sim	X(GFP)	X(Gal4)
mau Poxn posterior lobe enhancer	mau	X(GFP)	X(Gal4)
sim poxn rescue	sim		X
mau poxn rescue	mau		X

Table 4.4. Transgenic lines analyzed.

For each construct generated, the species from which it was cloned is listed, as well as the insertion site, and type of reporter (GFP or GAL4).

5.0 ASSESSING CONSTRAINTS ON THE PATH OF REGULATORY EVOLUTION

This chapter was published online in *Proceedings of the Royal Society B* on 11 November 2013; DOI: 10.1098/rstb.2013.0026. Mark Rebeiz helped design the project, analyze experimental data, and helped write the paper.

5.1 INTRODUCTION

Evolution often proceeds through the accumulation of numerous mutations that collectively generate meaningful phenotypic outcomes (Frankel et al., 2011; Geffeney, Fujimoto, Brodie, & Ruben, 2005; Nachman, Hoekstra, & D'Agostino, 2003; Neitz, Neitz, & Jacobs, 1991; Rebeiz et al., 2009). The order in which such changes are introduced may differ substantially in the functional consequences of intermediates. Although much attention has been focused on the path by which proteins evolve (Bridgham, Carroll, & Thornton, 2006; Ortlund, Bridgham, Redinbo, & Thornton, 2007; Weinreich, Delaney, Depristo, & Hartl, 2006), the constraints and complications that arise during the multistep evolution of non-coding transcriptional activating sequences (enhancers) are less understood. Moreover, regulatory DNA has become increasingly appreciated as a major source of phenotypically relevant variation, particularly contributing to the evolution of morphology (Chan et al., 2010; Cretekos et al., 2008; Gompel et al., 2005; McGregor et al., 2007; Rebeiz et al., 2009).

Although enhancers are frequently conserved (Hardison, 2000; Loots et al., 2000; Peterson et al., 2009; Rebeiz, Castro, Liu, Yue, & Posakony, 2012; Woolfe et al., 2005), they often diverge more rapidly than protein-coding sequences (Blow et al., 2010; Bradley et al., 2010; Dermitzakis & Clark, 2002). This is in part due to the constraints that the triplet amino acid code imposes on protein-coding DNA. Enhancers contain assemblages of docking sites for transcription factors that collectively influence the initiation rate of transcription (Levine, 2010). Great amounts of variation can be observed in the presence, spacing, and sequence of transcription factor binding sites within and between species, often resulting in regulatory sequences that maintain function despite extreme sequence variation (Balhoff & Wray, 2005; Ludwig, Bergman, Patel, & Kreitman, 2000; Swanson, Evans, & Barolo, 2010). Therefore, in order to assess how an enhancer might accumulate a number of functionally relevant changes, one must look to either slowly evolving regions, or at differences that have arisen over short evolutionary periods.

5.1.1 Possible Constraints on the Evolution of Regulatory DNA

There are several possible constraints that may disfavor certain mutational paths of regulatory sequences. These may include the preservation and improvement of the evolving activity, the maintenance of pre-existing functions, and the context dependence of mutations (Frankel et al., 2011; Rockman & Wray, 2002). In the case of adaptive evolution driven by constant directional selection, it is generally accepted that an evolving protein-coding or regulatory DNA must improve, or not diminish, the fitness of the organism with each step (Orr, 2005). This constant refinement of a derived activity can be constrained by epistatic and pleiotropic interactions.

Pleiotropy, the effect of a single mutation on multiple traits, appears to be a major constraint on evolutionary paths (Stern & Orgogozo, 2008). Although the pleiotropic consequences of mutations to regulatory DNA are predicted to be milder compared to protein-coding regions (Carroll, 2008; Stern, 2000), individual mutations may nonetheless lead to context-specific pleiotropic consequences. These may involve effects on other expression patterns of the gene in question, alteration of the regulation of adjacent genes (Cande et al., 2009), or the occurrence of unwanted ectopic expression. However, these pleiotropic effects may be circumvented by transitions that include epistatic interactions.

Epistasis, the dependence of a mutation's effect on the genetic background, could cause a path to be less favored compared to other paths that successively increase expression. An extreme case of epistasis, sign epistasis (Weinreich, Watson, & Chao, 2005), generates opposite effects of a mutation in different backgrounds. In a system under strong positive directional selection, where each step must increase an activity, paths that exhibit sign epistasis would be strongly disfavored. During the course of protein evolution, sign epistasis often restricts evolutionary trajectories that pass through structurally unstable intermediate states (Ortlund et al., 2007; Weinreich et al., 2006). However, regulatory sequences have been posited to be less susceptible to such destabilizing mutations (Rebeiz et al., 2009).

Robustness, the generation of reproducible outcomes in response to a highly varied environment has been a topic of much recent interest in the field of regulatory biology. For example, the maintenance of robustness has been cited as a cause for the existence of "shadow enhancers" (Frankel et al., 2010; Perry, Boettiger, Bothma, & Levine, 2010), the phenomenon that often multiple enhancers exist for a similar activity in the same gene (Hong et al., 2008). In two separate instances, the removal of a shadow enhancer, while maintaining the other copy has

caused a decrease in robustness: animals lacking the "shadow" copy show greater variability in phenotype when grown at differing temperatures or in differing genetic backgrounds (Frankel et al., 2010; Perry et al., 2010). It has also been shown that apparently redundant binding sites within a single enhancer may be required to foster robustness (Ludwig, Kittler, White, & Kreitman, 2011). It is generally assumed that the establishment of robustness is a common step during an enhancer's evolution. However, we currently lack examples that demonstrate an enhancer evolving from a less robust state into a more robust state.

5.1.2 A Model for Studying the Path of Regulatory Evolution

While the constraints on a regulatory sequence's evolution can be easily imagined, we currently lack fundamental knowledge of what is possible during an enhancer's path of evolution. How pervasive is epistasis? What kinds of epistatic interactions exist? When and how can robustness evolve? What other unexpected constraints on enhancer evolution exist? Given the prevalence and rapidity of regulatory DNA evolution, the identification of forces constraining evolutionary paths represents an important step in understanding how regulatory sequences acquire altered functions.

Previous studies in the lab elucidated the origins of a newly evolved enhancer activity that arose in the *Nep1* gene of *D. santomea* (Rebeiz et al., 2011). Optic lobe expression of *Nep1* in lamina precursors (Figure 5.1B-C) is unique to the *D. santomea* visual system. This novel expression pattern is encoded by a 680bp enhancer element embedded in the first intron of the *Nep1* gene (Figure 5.1A,D). The novel optic lobe activity of *D. santomea Nep1* overlaps several other enhancer regions in the intron, suggesting that perhaps this activity sprouted out of a pre-existing adjacent enhancer (Figure 5.1A). In a series of mutant reporters, we determined that the

novel optic lobe activity depends upon short stretches of nucleotides required for full activity of other ancestral overlapping activities in the retinal field and central nervous system (CNS). Further, by sequencing this segment from multiple (≥14) isofemale lines of *D. yakuba, D. santomea,* and the closest outgroup *D. teissieri*, we found that the *D. santomea* optic lobe enhancer differs from the *D. yakuba/santomea* ancestor by just four fixed mutations. In an *in vivo* reporter assay, we found that reversion of each of these mutations in the context of the *D. santomea* enhancer led to a significant reduction in activity. By reverting all four of these mutations simultaneously, we tested the activity of the resurrected *D. yakuba/santomea* ancestral enhancer. We found that this enhancer had a weak activity in the optic lobe, suggesting that this was the starting point for the strong, derived optic lobe expression of *D. santomea Nep1*. Although it is uncertain whether the changes at *Nep1* were adaptive, its novel optic lobe activity is unique in that it is an experimentally tractable example in which a short path of mutations leads to greatly increased enhancer activity.

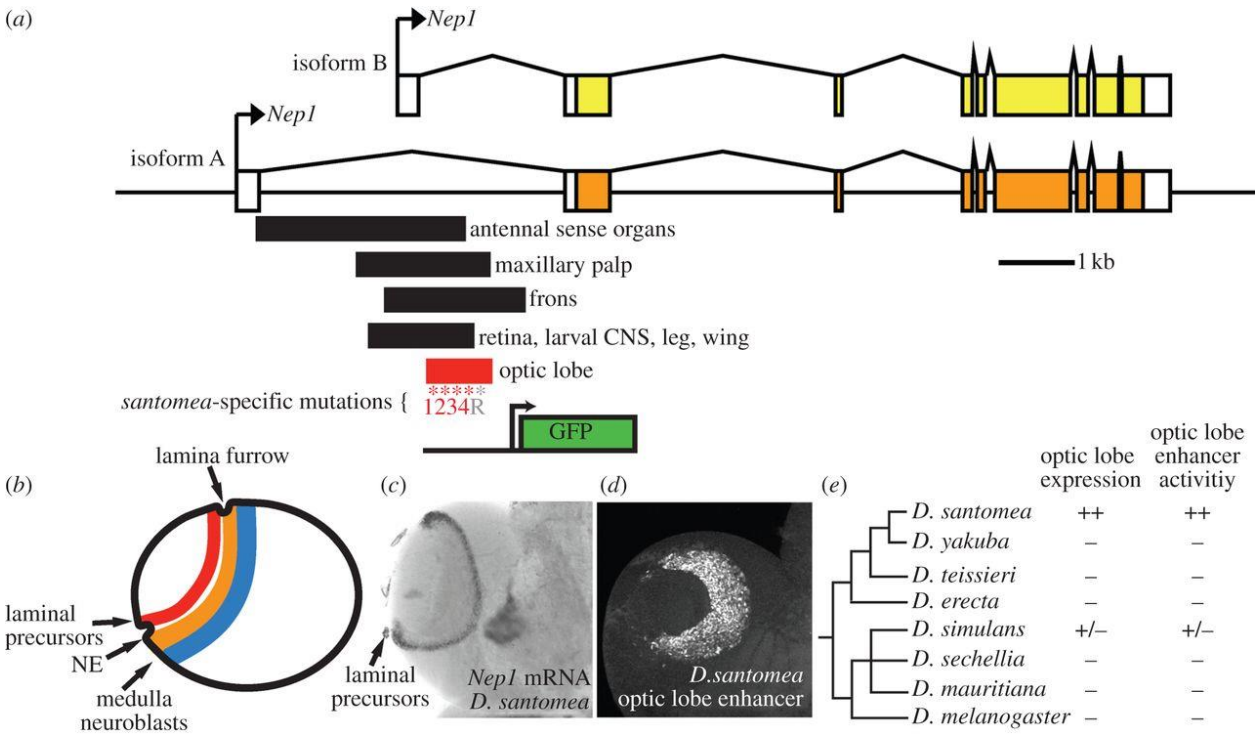

Figure 5.1. A recently evolved expression pattern of the *Nep1* gene depends upon four mutations fixed in *D. santomea*.

(A) The first *Nep1* intron contains several overlapping transcriptional enhancers, including an optic lobe enhancer novel to *D. santomea*. The *D. santomea* enhancer differs from the derived *D. yakuba/santomea* ancestral optic lobe enhancer by four mutations, marked by red asterisks. The position of a polymorphic repeat expansion in modern-day *D. santomea* is marked by a grey asterisk. The extent of a GFP reporter construct used in this study is denoted. (B) *Nep1* optic lobe activity occurs in laminal precursors, a transition state that occurs as neuroepithelial (NE) cells migrate past the lamina furrow and become lamina neurons. NE cells that migrate away from the furrow transition to medulla neuroblasts. (C) *In situ* hybridization of the *D. santomea* third instar larval optic lobe with a *Nep1* riboprobe reveals *Nep1* expression in lamia precursor cells (arrow) and the mushroom body. (D) The *D. santomea* optical lobe enhancer reporter construct drives expression in lamina neurons. (E) Phylogeny of *D. yakuba* and *D. melanogaster* clades referencing optic lobe expression assayed via *in situ* hybridization and by species-specific reporter constructs. (+) denotes strong expression, (-) denote absence of expression, (+/-) denotes weak expression (Rebeiz et al., 2011).

Here, we use the recently evolved optic lobe activity of *Nep1* to assess constraints that may influence the path of an enhancer's evolution. From the starting point of the reconstructed *D. yakuba/santomea* ancestral enhancer sequence, we tested each possible evolutionary intermediate in an *in vivo* reporter assay in order to identify ways in which the order of introduction may be restricted. First, we observe sign epistasis: introduction of certain mutations can increase or decrease activity depending on the mutational trajectory. Further, we noted that some paths modulate the activities of an overlapping enhancer, which could influence fitness. Finally, we observe paths that progress through intermediates with strong ectopic activity that manifest under chronic temperature stress. These results provide empirical evidence of the types of constraints that are likely to influence the ordering of mutations that are acceptable during the diversification of regulatory sequences under persistent, directional selection for increased expression.

5.2 RESULTS

To detect possible constraints on the evolutionary ordering of regulatory sequence mutations, we constructed mutant versions of an 1176 bp non-coding DNA segment containing the *D. santomea Nep1* optic lobe enhancer (Rebeiz et al., 2011). Working with four mutations that were fixed in the *D. santomea* lineage, we generated and tested all possible combinations of these mutations in the context of this fragment (Figure 5.2A). These versions represent all possible evolutionary intermediates along the trajectory from the reconstructed ancestor of *D. yakuba* and *D. santomea* to modern day *D. santomea.* By inserting these constructs into the same genomic position as in our previous study, we were able to control for positional effects on reporter

activity. Previously, the reconstructed ancestor was engineered to remove a repeat expansion that is polymorphic in *D. santomea* (Rebeiz et al., 2011) (Figure 5.1A, "R"). To recreate intermediates from the *yakuba/santomea* ancestor to the modern day *D. santomea*, we added this repeat, which represents the most common allele in the sample (4/14 sequences). Doing so resulted in a 22% increase in expression from 30% to 52% relative to the modern-day *D. santomea* construct (Figure 5.1A). Thus, this highly divergent region that is mostly composed of unique alleles in our sequence sample (Rebeiz et al., 2011) influences the optic lobe activity.

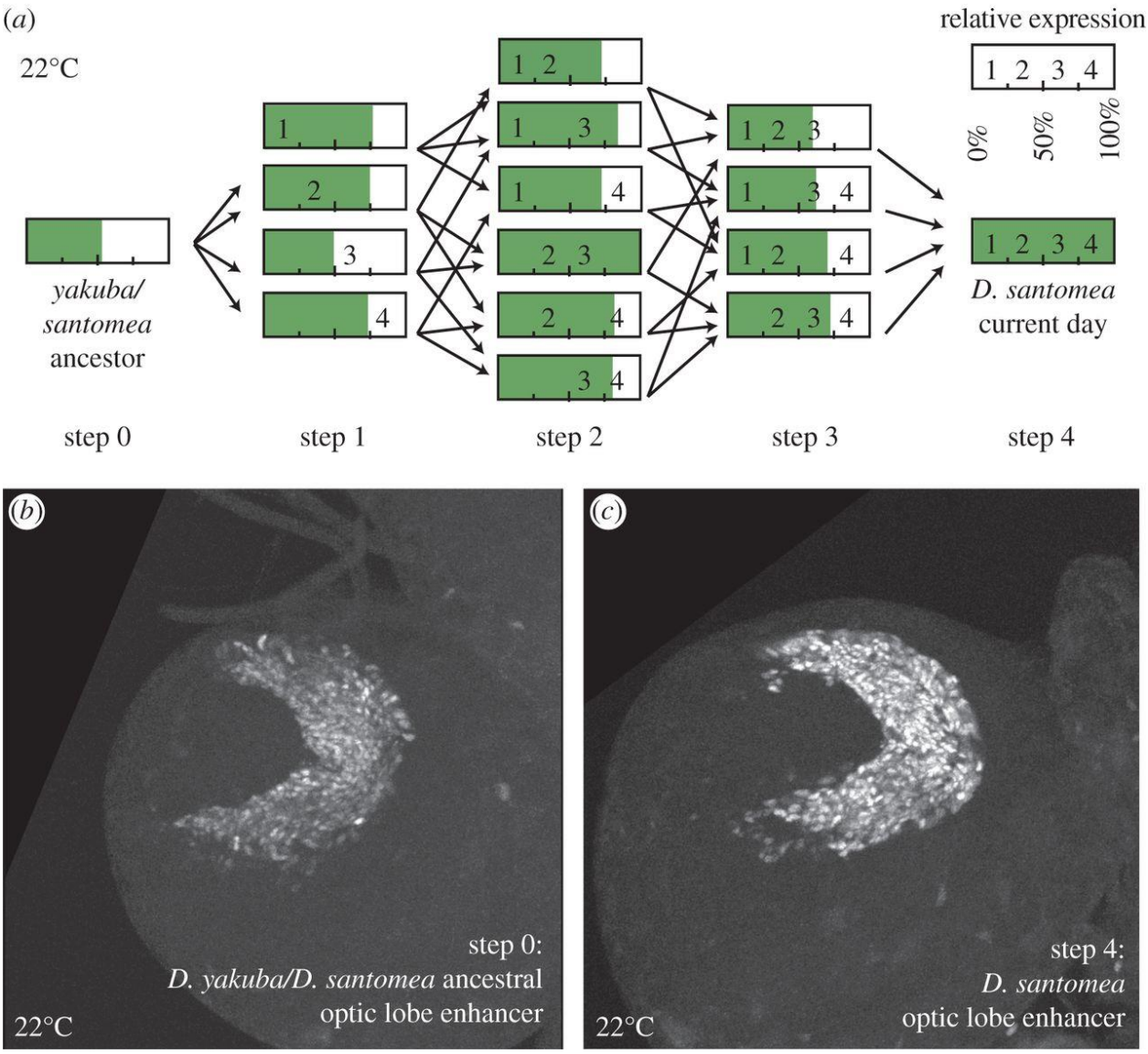

Figure 5.2. Map of the effects of mutational paths on enhancer activity in the optic lobe.

(A) Schematic of intermediates between the *D. yakuba/santomea* ancestral optic lobe enhancer and the *D. santomea* modern day optic lobe enhancer illustrates possible evolutionary pathways. Each bar represents a reporter construct. Numbers within each bar denote the presence of one or more mutations on the path to modern day *D. santomea*. The green shading represents reporter expression quantified from the lamina neurons of third instar larval brains grown at 22°C, relative to the *D. santomea* construct. (B) Optic lobe of *D. yakuba/santomea* ancestral optic lobe enhancer reporter construct. (C) optic lobe of *D. santomea* modern day optic lobe reporter construct.

Overall, the activity level of these intermediates varied from low expression similar to the ancestral enhancer (52% of modern day *D. santomea*) to high activity that resembled the modern day *D. santomea* construct (Table 5.1, Figure 5.2A). The dataset was analyzed by one-way ANOVA followed by pairwise comparisons, using post-hoc Tukey's HSD tests ($\alpha = 0.05$), which statistically accounts and corrects for multiple, simultaneous comparisons. Below, we report observed constraints that may restrict the path by which this enhancer could have accumulated these four fixed differences.

Construct	N	Relative Activity	SEM	Class*				
Ancestor	36	52.5%	2.5%				D	E
Ancestor FW1	31	78.8%	3.5%			C		
Ancestor FW2	38	76.6%	2.8%			C		
Ancestor FW3	27	49.2%	2.6%					E
Ancestor FW4	51	72.4%	2.0%			C		
Double Mut 1-2	20	71.5%	4.4%			C	D	E
Double Mut 1-3	16	84.2%	6.1%	A	B	C		
Double Mut 1-4	14	72.7%	6.8%			C	D	E
Double Mut 2-3	55	102.6%	3.2%	A				
Double Mut 2-4	42	81.2%	4.6%			C		
Double Mut 3-4	20	79.1%	4.4%		B	C		
Triple Mut 2-3-4	16	73.0%	7.6%			C	D	E
Triple Mut 1-3-4	16	65.5%	8.1%			C	D	E
Triple Mut 1-2-4	22	72.3%	4.0%			C	D	
Triple Mut 1-2-3	59	55.8%	2.4%				D	E
Santomea	37	100.0%	4.9%	A	B			

Table 5.1. Relative expression levels of *Nep1* optic lobe enhancer intermediates

* Mutant constructs connected by the same letter are not significantly different (p >.05). SEM: standard error of the mean.

5.2.1 Epistatic Effects

Upon reconstructing the possible ways that the four fixed mutations could have accumulated, we noticed non-additive interactions between the mutations 2 and 3 (Figure 5.3). Although these mutations had significant effects when removed from *D. santomea* (Table 5.1, Triple Mut 1-3-4, Triple Mut 1-2-4), mutation 3 showed no significant effect when introduced in the background of the ancestral construct (Figure 5.3, Ancestor Forward 3). However, when mutation 2 is added to mutation 3 (Figure 5.3, Double Mut 2-3), a ≈2-fold increase in enhancer activity was observed. Thus, the presence of mutation 2 is required for the effect of mutation 3. Several additional intermediate steps similarly represented "lateral moves" in which a significant increase in activity was not detected. Indeed, of 24 possible paths connecting the ancestor to modern day *D. santomea*, 22 contained steps that did not significantly increase activity in uncorrected pairwise t-tests ($\alpha = 0.05$). Thus, although each mutation is required for modern-day *D. santomea* activity levels, nearly every path includes a transition that does not noticeably increase expression. Moreover, many paths, including the remaining 2 of 24 exhibit steps resulting in a significant decrease in expression (discussed further below).

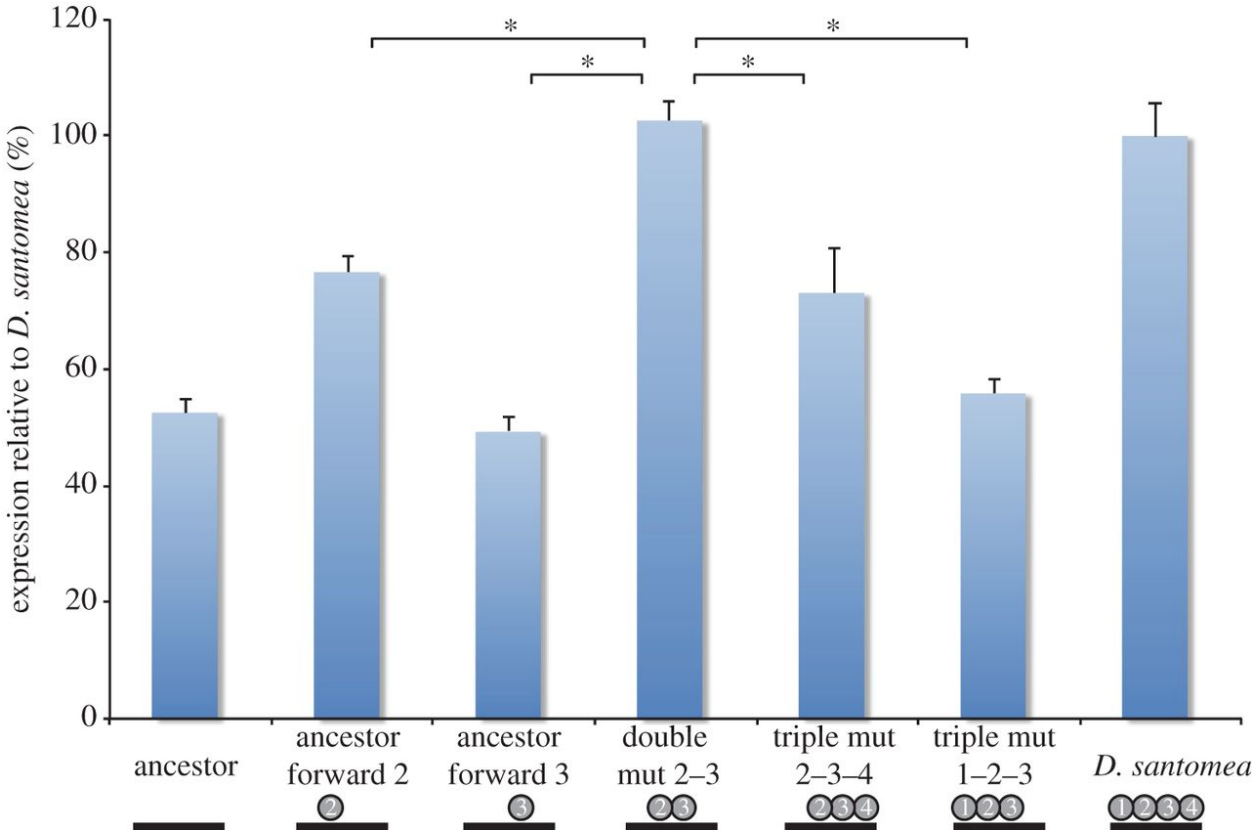

Figure 5.3. Epistatic non-additive interactions and sign epistasis between mutations mark the path of the *Nep1* enhancer's evolution.

Graph comparing the mean activity of optic lobe enhancers that contain or lack mutations 2 or 3 reveal epistatic non-additive interactions. While addition of mutation 3 to the ancestral construct does not result in a significantly different activity increase (Ancestor Forward 3), addition of mutation 2 to this construct results in a drastic increase in activity (Double Mut 2-3). Sign epistasis is revealed upon the introduction of 1 (Triple Mut 1-2-3) or 4 (Triple Mut 2-3-4) into this background, which both decrease activity. All activity values are normalized to modern day *D. santomea* levels. Error bars show standard error of the mean. Asterisks denote significant differences (Tukey's HSD test, $\alpha = 0.05$).

5.2.2 Sign Epistasis

In addition to epistatic interactions, we also observed that several mutational paths involved significant sign epistasis (Figure 5.3). For example, the intermediate that combines mutations 2 and 3 has a high activity, 103% of modern day *D. santomea* (Figure 5.3). Subsequent addition of mutations 1 or 4 leads to a 47% or 30% reduction in activity, respectively (Figure 5.3 "Triple Mut 2-3-4", "Triple Mut 1-2-3"). Of the 24 possible paths, a full six of these include steps that show significant sign epistasis ($p < 0.05$, Tukey's HSD). Six additional paths suggested sign epistasis, with transitions that were detected as significant decreases in an uncorrected paired t-test ($p < 0.05$). Thus, a full half of the possible trajectories connecting the *D. yakuba/santomea* ancestor to modern day *D. santomea* included significant or suggestive sign-epistasis.

5.2.3 Ectopic Expression

During our analysis of the possible evolutionary trajectories, we noticed ectopic expression in some paths, manifesting in a zone of the medulla, adjacent to the laminal precursors that express *Nep1* in *D. santomea* (Figure 5.4B, arrow). This ectopic activity was particularly noticeable when lines were reared under chronic temperature stress at 30°C. The resurrected ancestor of *D. yakuba* and *D. santomea* had a fairly high level of ectopic expression in this location (Figure 5.4B), while *D. santomea* exhibited little to no ectopic expression (Figure 5.4C). Measuring the ectopic expression of different trajectories at 30°C, we noted that distinct paths increased or decreased this ectopic activity to differing extents (Figure 5.4A). For example, addition of mutation 4 to the ancestor led to enhanced ectopic expression (≈1.5-fold increase, $p < 0.05$, uncorrected paired t-test), while introduction of mutation 2 or 3 to Ancestor Forward 4

completely ablates it. Most paths that didn't include the early addition of mutation 4 exhibited the general trend of reducing ectopic activity (Figure 5.4A).

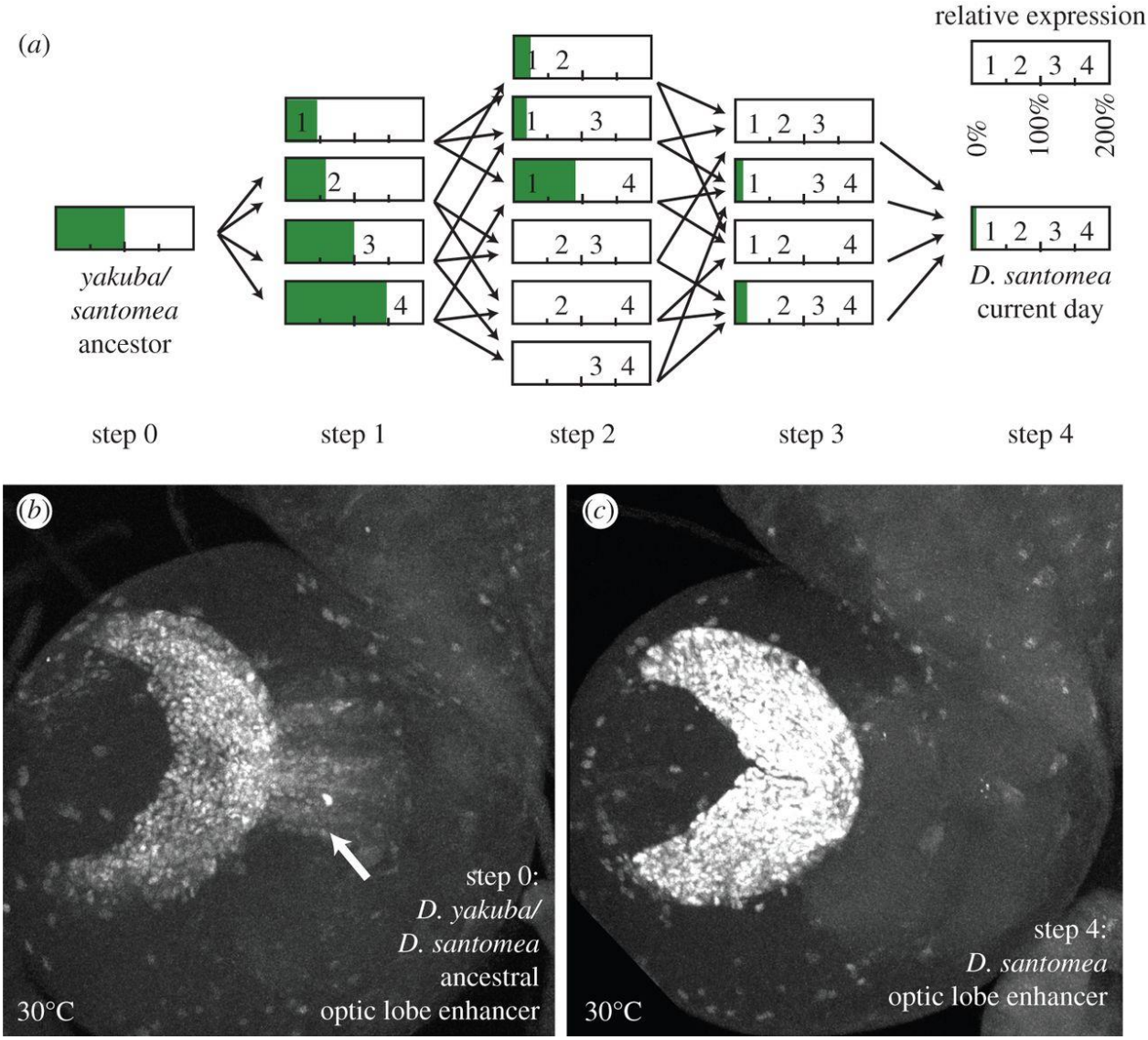

Figure 5.4. The effects of mutational path on ectopic expression.

(A) Schematic of intermediates between the *D. yakuba/santomea* ancestral optic lobe enhancer and the *D. santomea* modern day optic lobe enhancer illustrating possible evolutionary pathways. Each bar represents a reporter construct, and numbers denote which *D. santomea*-specific mutations are present. Green shading represents reporter expression quantified from the medullar neuroblasts of third instar larval brains grown at 30°C, normalized to the level of the *D. yakuba/santomea* ancestral construct (B) Optic lobe of *D. yakuba/santomea* ancestral optic lobe

enhancer reporter construct grown at 30°C reveals ectopic expression in a region of the medulla (arrow), (C) optic lobe of *D. santomea* modern day optic lobe reporter construct grown at 30°C lacks this ectopic activity.

5.2.4 Effects on Overlapping Activities

In addition to pleiotropic activation of the reporter in ectopic locations, we observed intermediates that had effects on a different tissue where *Nep1* is deployed: the larval CNS (Figure 5.5A). Larval CNS expression exhibited by the ancestral construct (Figure 5.5B) is absent in the modern day *D. santomea* construct (Figure 5.5D), and is increased in an intermediate construct, Ancestor Forward 4 (Figure 5.5C). These results suggest that different trajectories would appear to modulate existing CNS activity of *Nep1*, and larval CNS activity is reduced during the evolution of the optical lobe enhancer from the *D. yakuba/santomea* ancestral state. We conclude the path of the *Nep1* optic lobe enhancer's evolution can alter overlapping endogenous functions of the *Nep1* gene that may impact fitness.

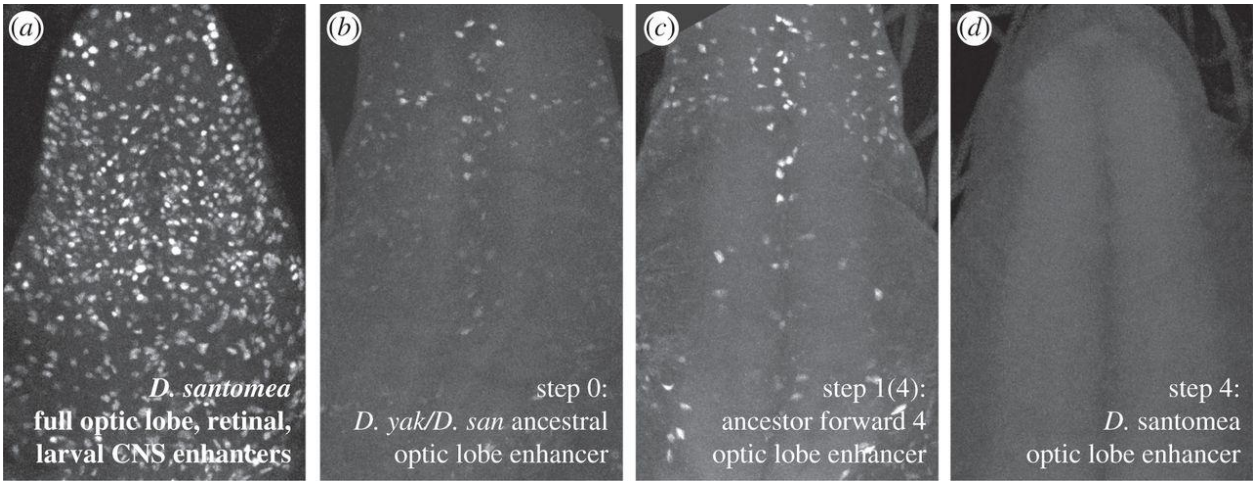

Figure 5.5. Individual paths differentially contribute activity to a known expression pattern of *Nep1*.

(A) Expression of a *Nep1 r*eporter construct containing the retinal field, larval CNS and optic lobe enhancers in the ventral ganglion. (B-D) The *D. yakuba/santomea* ancestral optic lobe enhancer reporter is expressed in a small proportion of the ventral ganglion (B) while the *D. santomea* optic lobe enhancer drives no expression (D). The Ancestor Forward 4 construct drives increased expression in the ventral ganglion (C).

5.3 DISCUSSION

Here, we have examined several factors that may commonly restrict a regulatory sequence's path of evolution. By generating and testing a comprehensive set of all possible evolutionary intermediates in an *in vivo* assay, we explored the biological pitfalls of individual mutational paths, and compared their merits. Although each of the four mutations we characterized increased activity in at least one setting, every single path included non-additive or sign-epistatic legs along the journey to the modern-day *D. santomea Nep1* enhancer. Above and beyond the sign and magnitude of expression differences between intermediates, our findings suggest that not all paths are equal in terms of pleiotropic effects on pre-existing and ectopic activities. Nevertheless, no combination of mutations caused the enhancer to fail utterly. Though we cannot

comment on the biological significance of *Nep1* expression in the optic lobe, or what negative fitness consequences would result from paths that induced pleiotropic effects, the constraints that are revealed by our study illuminate what is possible for an evolving regulatory sequence. Indeed, some of the constraints we examined would pertain to both adaptively and neutrally evolving enhancers. These findings provide a more nuanced view of the complexities associated with evolving increased enhancer activity.

5.3.1 Epistatic Interaction and Enhancer Information Processing Mechanisms

Under a model of persistent directional selection, epistasis is predicted to constrain potential paths of evolution. Although our initial experiments with the *Nep1* enhancer of *D. santomea* suggested that each of the four fixed mutations are required to generate the full activity of the *D. santomea* enhancer (Rebeiz et al., 2011), our reconstruction of all possible paths revealed how the process of introducing these mutations in sequence was not straightforward. Indeed, each of the four mutations had contexts in which their addition had no effect on expression level. Mutation 3 presents a very clear case of cooperative interaction (Figure 5.3), as it only increased activity in a limited number of contexts (Figure 5.2). Moreover, the polymorphic repeat expansion (Figure 5.1A, “R”) introduces additional epistatic interactions (Figure 5.6), illustrating how polymorphisms could further complicate the interpretation of mutational effects.

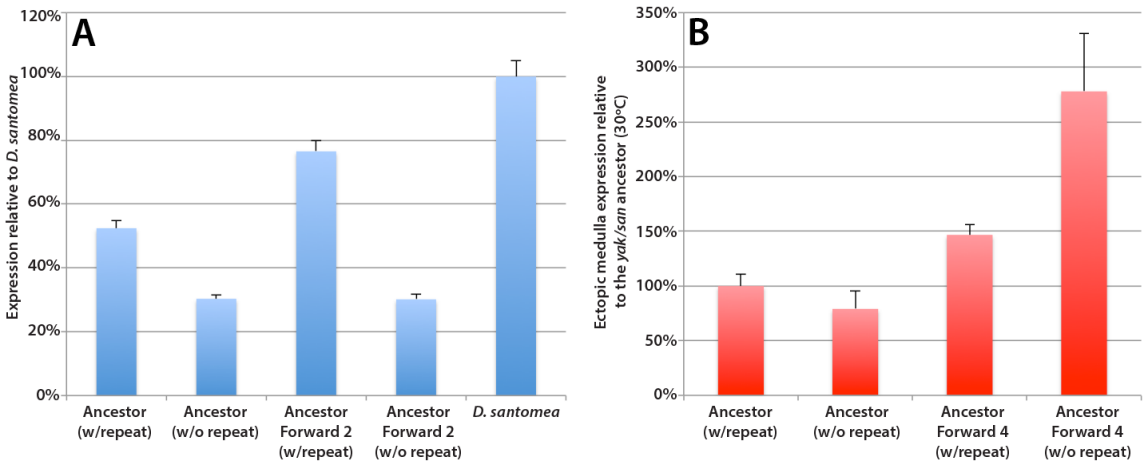

Figure 5.6. A high frequency polymorphism in modern-day *D. santomea* affects enhancer activity.

(A) Graph depicting activities of the *D. yakuba/santomea* ancestral enhancer with and without the *D. santomea*-specific polymorphic repeat. Values are expressed relative to the modern-day *D. santomea* construct. Constructs bearing the forward mutation 2 with and without the repeat are also compared, demonstrating that mutation 2 depends on this polymorphic region to increase activity in this context. (B) Graph depicting ectopic expression in the medulla in animals reared at 30°C. Expression relative to the *D. yakuba/santomea* ancestor containing the repeat region are depicted. In this instance this polymorphic region shows context specificity in enhancing ectopic expression.

In a now highly influential review, Arnosti and Kulkarni put forward two contrasting models of how enhancers process information: billboards and enhanceosomes (Arnosti & Kulkarni, 2005). In the enhanceosome model, the enhancer DNA acts as a scaffold to form a higher-order conformation of interacting proteins. Such a model is supported by the precise requirement for the presence and spacing of all of the binding sites in the enhancer to activate transcription (Thanos & Maniatis, 1995). In contrast, the billboard model suggests that spacing and cooperative interaction of binding sites is minimal, and that the net output of such an enhancer is the collective interpretation of positive and negative inputs. Although it is well-

recognized that enhancers may incorporate aspects of enhanceosome and billboard architecture simultaneously (Arnosti & Kulkarni, 2005), these two contrasting models of enhancer action are predicted to differ in the flexibility of their evolutionary paths. A billboard enhancer that is evolving new binding sites would be predicted to be unconstrained by epistatic effects, while an enhancer that follows the enhanceosome model would have many (if not all) paths that include epistasis.

Considering our data in light of the enhanceosome and billboard models, we suggest that the derived activity of the *D. santomea* optic lobe enhancer of *Nep1* likely represents a combination of both. The widespread epistatic effects we observe are consistent with the evolution of binding sites for proteins that interact physically, as expected of an enhanceosome. However, in none of the intermediates is expression completely lost, or reduced below the level observed for the ancestor. Thus, the aspects of the optic lobe enhancer are consistent with a billboard architecture as well.

5.3.2 The Prevalence and Possible Mechanisms of Sign Epistasis

The prevalence of sign epistasis in protein coding sequences is often attributed to tradeoffs between thermodynamic stability and the evolution of new functions (Ortlund et al., 2007; Weinreich et al., 2006). In the case of TEM β-lactamase, Weinreich and colleagues observed sign epistasis between a mutation that increases antibiotic hydrolysis while concurrently reducing its stability, with a second mutation that increases thermodynamic stability, but slightly reduces activity. For several of the paths, introduction of the stabilizing mutation was deleterious in the absence of the activity-increasing mutation (Weinreich et al., 2006). In contrast, Ortlund and colleagues found that mutations which increased thermodynamic stability were required

before major function-altering mutations could evolve in the vertebrate glucocorticoid receptor. In this case, the mutations to the binding pocket of the protein were so dramatic that unstable intermediates would form in the absence of these permissive mutations (Ortlund et al., 2007). As thermodynamic instability represents a dead-end for an evolving protein, sign epistasis is expected to be a rigid constraint during the path of coding sequence evolution.

An unexpected finding of this work was the frequency of sign epistasis among the reconstructed evolutionary trajectories of a regulatory sequence. Each and every mutation had at least one context in which its introduction would decrease the expression level from a previous step (Figure 5.7). There are several possible explanations for how mutations to an enhancer could generate opposite effects on expression. If the mutations generate new, or higher affinity binding sites for a particular factor, it is possible that the context of adjacent transcription factor binding events could influence the recruitment of activating or repressive complexes. Alternately, the evolution of a strong binding site may cause other factors previously bound to the region to be displaced. If a cooperative interaction between two factors is evolving, the intermediate step in which just one factor is present may cause a reduction in activity, simply due the displacement of a protein that was previously contributing to the activity. Future elucidation of the transcription factors that comprise the *Nep1* optic lobe enhancer will allow us to distinguish these and other competing models of how regulatory mutations interact cooperatively and antagonistically.

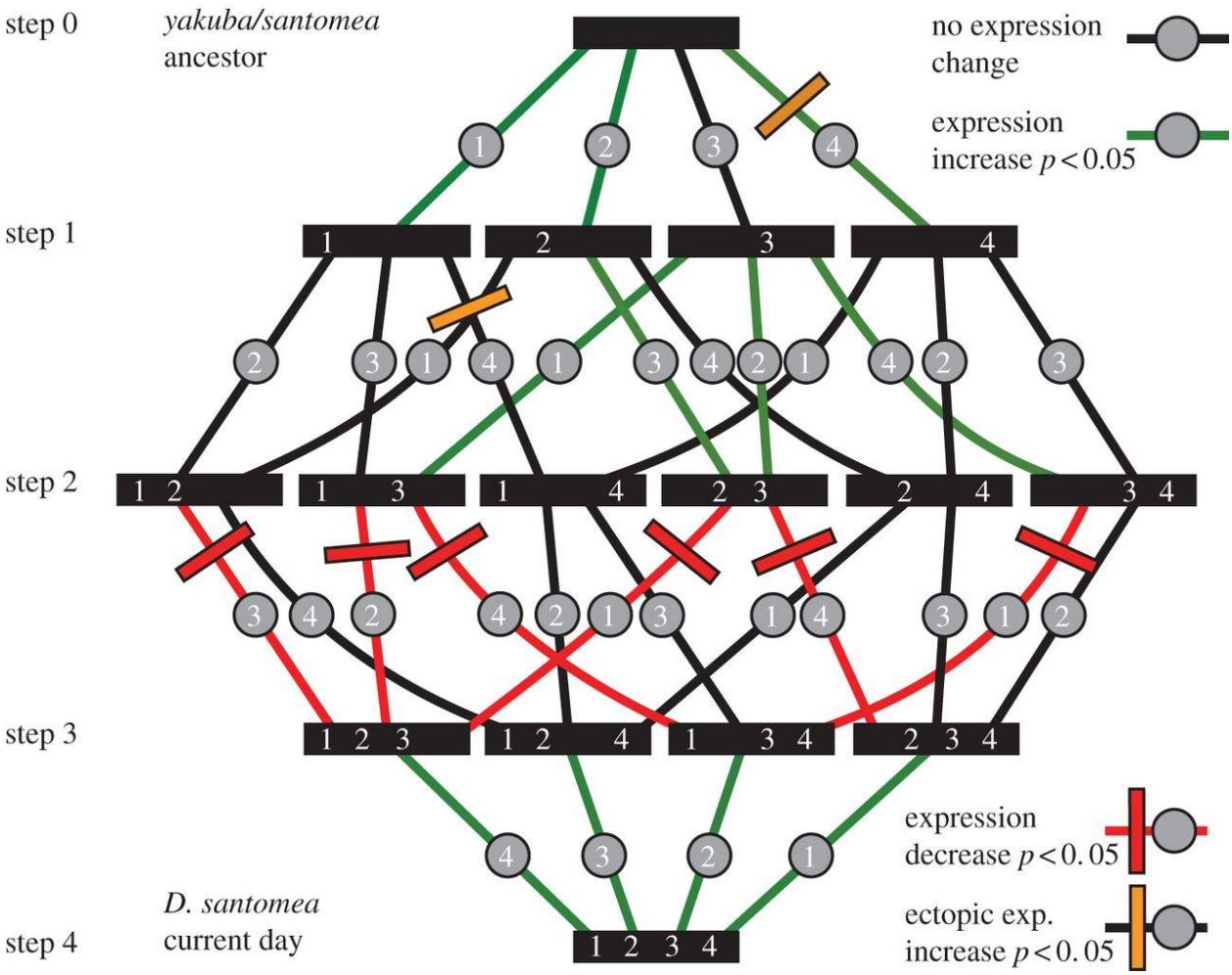

Figure 5.7. Summary of constraints restricting the path of evolution of the *Nep1* enhancer.

This schematic illustrates the constraints that may guide the path of evolution of the *Nep1* enhancer. The black bars with gray numbers represent possible mutational intermediates between the *D.yakuba/D.santomea* ancestral enhancer and the modern day *D. santomea* enhancer. Lines represent possible evolutionary pathways between the ancestral and the modern day enhancer: green paths represent a significant increase in expression (Tukey's HSD $\alpha < 0.05$), red paths represent a significant or suggestive decrease in expression ($p < 0.05$) and black paths represent an insignificant change in expression. Colored bars that cross individual routes represent possible constraints that may reduce the viability of a given evolutionary pathway: red bars are on paths with a suggestive decrease in activity ($p < 0.05$), orange bars are on paths with a suggestive increase in ectopic activity ($p < 0.05$).

5.3.3 Co-option of Existing Activities – Opportunities for Novelty and Pleiotropy

Previous work has posited several mechanisms by which new regulatory activities may arise (Britten, 1996; de Souza, Franchini, & Rubinstein, 2013; Eichenlaub & Ettwiller, 2011; Rebeiz et al., 2011). Although evidence exists for several of the possible mechanisms (Cande et al., 2009; Chung et al., 2007; Eichenlaub & Ettwiller, 2011; Gompel et al., 2005), the *Nep1* optical lobe enhancer represents an example of co-option of a pre-existing regulatory activity (Rebeiz et al., 2011). The re-use of pre-existing architecture through co-option offers many advantages over the stepwise evolution of enhancers *de novo* by allowing complex regulatory schemes to be built in fewer evolutionary steps. However, in both adaptive and neutral evolutionary contexts, it also poses distinct challenges. New activities that evolve in the middle of existing enhancers run the risk of altering the activity of those enhancers, making the evolutionary path susceptible to pleiotropic effects. This characteristic can be seen as a structural constraint that is unique to the origination of new enhancers by co-option. Our study uncovered evidence that intermediates during the evolution of the optic lobe enhancer drive differing levels of expression in tissue regulated by a pre-existing overlapping enhancer. While it is uncertain whether the full regulatory region is able to buffer the pleiotopic effects incurred by the evolution of optic lobe activity, our data illustrates a constraint that may govern the modification of a co-opted enhancer.

5.3.4 The Pleiotropic Effects of Ectopic Expression

Mis-expression of a gene can be catastrophic. This is evidenced by the widespread incidence of such effects in genetic disorders and disease (Shastry, 1995). Overexpression is a widely used tool in genetic research, precisely because it often produces phenotypes that are not visible

during loss of function studies (Prelich, 2012). Evolutionary paths that lead to ectopic expression will be instantly evaluated by selection as these routes would convey dominant effects on expression. Thus, the *Nep1* case raises the possibility that the path of enhancer evolution may be commonly restricted to paths that eliminate ectopic expression.

Enhancers harbor an enormous potential to generate ectopic activities. The transcription factors that activate enhancers are deployed repeatedly in many locations during development (Wray et al., 2003). Although the binding of an upstream transcription factor could lead to activation in a multitude of tissues, "combinatorial logic" is thought to restrict an enhancer's activity to one or a few developmental contexts (Levine, 2010). Nevertheless, two recent examples demonstrate how rearrangement of existing binding sites in an enhancer can generate novel ectopic activities (Liu & Posakony, 2012; Swanson et al., 2010). In a striking example of an enhancer's potential to generate ectopic activity, Liu and Posakony demonstrated how the same combination of transcription factors mediate expression of distinct target genes in two separate Notch-responsive settings during *Drosophila* development (Liu & Posakony, 2012). A simple shift in the position of a POU-HD binding site within the *Enhancer of split mα* enhancer was sufficient to cause weak expression in additional Notch responsive settings. Thus, although a combination of binding sites may generate expression in multiple tissues, their relative positioning and orientation may be instrumental in controlling the enhancer's specificity. Our results resonate with these studies in that the order in which mutations are introduced can influence the degree to which expression is observed in ectopic locations.

5.4 METHODS

5.4.1 Transgenic Constructs

Mutated versions of the *Nep1* enhancer fragment were produced by overlap extension PCR using primer sequences described previously (Rebeiz et al., 2011). Constructs differ only at the noted sites; the entire sequence of each construct was confirmed via sequencing. PCR products were cloned into the S3aG transgenesis plasmid (Williams et al., 2008) using *Asc I* and *Sbf I* sites. S3aG contains a multi-cloning site upstream of a basal promoter driving enhanced nuclear GFP derived from the pH-Stinger series of vectors (Barolo, Carver, & Posakony, 2000), as well as a donor attB site for site-specific insertion into the *Drosophila melanogaster* genome. Constructs were injected by Rainbow Transgenic Flies, Inc, (Camarillo, CA) into a ϕC31-integrase expressing line with an attP insertion on the second chromosome (51D) (Bischof et al., 2007). Independent transgenic lines were outcrossed to a *yellow-white* stock for two generations before the establishment of homozygous insertions.

5.4.2 Quantification of Reporter Activity

Late third instar female larvae for at least two lines were dissected in cold PBS, and fixed in 4% paraformaldehyde PBT (PBS + .1% Triton-X-100) solution for 30 minutes at room temperature. Samples were washed several times in PBT and then incubated in a 50% glycerol/PBT solution for 10 minutes before mounting on slides in glycerol mountant (80% Glycerol, .1M Tris, pH 8.0).

Mounted brains and imaginal discs were imaged on an Olympus Fluoview 1000 confocal microscope using standardized non-saturated settings. Maximum projections of reporter construct expressing brains were saved, and fluorescent intensity was quantified using the ImageJ software with the freehand selection tool. The region used for intensity measurements was chosen by making selections on duplicated images whose brightness was increased, and subsequently measuring these selections on un-manipulated images. Expression intensity was compared using one-way ANOVA in the JMP-pro software package (SAS Institute Inc., Cary, NC). Posthoc pairwise comparisons of sample means were carried out with the Tukey's HSD test.

6.0 DISCUSSION AND FUTURE DIRECTIONS

6.1 SUMMARY OF THE POSTERIOR LOBE NETWORK

6.1.1 Mapping the Components of the Posterior Lobe Gene Regulatory Network

In this dissertation, I identified several members of the novel posterior lobe gene regulatory network (GRN) by first studying the regulatory mechanisms governing the one gene previously known to be specifically necessary for the development of the posterior lobe, *Poxn* (Boll & Noll, 2002). Studying the ancestry of the *Poxn* posterior lobe enhancer (PLE), I discovered that it existed before the evolution of the posterior lobe and that it had been co-opted into the posterior lobe from the posterior spiracle. Identifying this shared circuit between the posterior lobe and posterior spiracle networks led to the identification of several further co-opted spiracle enhancers, including the *crb* PLE (Lovegrove et al., 2006), *en* "D" enhancer (Cheng et al., 2014), *Cad86C* PLE, *Gef64C* PLE, *eya* PLE, and *ems* downstream (DS) enhancer (Glassford et al., 2015). The genes of these co-opted enhancers, including the terminal differentiation genes *crb, Cad86c* and *Gef64C*, comprise the lower tier members of the posterior lobe network hierarchy (Figure 6.1).

Identifying novel upper tier members of the posterior lobe GRN required a different approach, as two of the top tier posterior spiracle network factors (*Abd-B* and *sal*), are not unique

to posterior lobed species, the third (*ems*) was co-opted directly from the posterior spiracle network (Glassford et al., 2015). The final top tier posterior spiracle network factor, the Jak/STAT signaling pathway ligand *upd*, does not possess a shared posterior spiracle and posterior lobe enhancer (Glassford et al., 2015), leading us to focus our study on its activity in the genitalia. Both *upd* and *Poxn* are expressed in two waves during genital development: the first wave is associated with the cleavage of the lateral plate from the clasper, and the second wave is a novel deployment in species with posterior lobes (Glassford et al., 2015). The first wave of *Poxn* is expressed before the expansion of *Delta* in the lateral plate and is likely upstream of the expanded posterior lobe associated activity of *Delta* (Figure 3.9), while the second wave of *Poxn* is most highly expressed after the expansion begins to recede and it likely downstream of *Delta* (Figure 2.4). Both of these waves of expression are directly regulated by Jak/STAT signaling, as the mutation of a conserved STAT binding site within a reporter of the *Poxn* posterior lobe enhancer eliminates both waves of activity (Figure 2.5). To understand the molecular basis of Jak/STAT signaling activity in the genitalia, transgenic reporters were created to screen the non-coding regulatory regions between the three Jak/STAT signaling ligands: *upd, upd2* and *upd3* (Sarah Smith, personal communication). Two genital enhancers of *upd* were identified: an enhancer called MA5 that initiates early in genital development, and an enhancer called MA8 that is expressed during posterior lobe development and also is active in the larval wing hinge and eye patterns of *upd* (Sarah Smith, personal communication). Further analysis found that the orthologous *upd* MA5 region from non-lobed species is active, suggesting that the MA5 enhancer is ancestral to the evolution of the posterior lobe, although it is expressed for a longer period of time than seen in non-lobed species (Sarah Smith, personal communication). Interestingly, the orthologous MA8 region from non-lobed species is not active, suggesting that

MA8 is also novel to posterior lobed species (Sarah Smith, personal communication). Also, MA8 enhancer activity is almost completely reduced by the expression of RNA hairpins targeting *Delta*, suggesting that MA8 is downstream of *Delta* (Sarah Smith, personal communication). Neither of these genes are associated with completely novel regulatory activity, however, as the first wave of *Poxn* is driven by an ancestral *Poxn* posterior lobe enhancer and the early activity of *upd* is likely driven by the ancestral MA5 enhancer (Figure 6.1). As both of these genes and their enhancers existed before the evolution of the posterior lobe, they are unlikely to be the novel upstream regulators responsible for the expansion of *Delta* in posterior lobed species. It is possible, however, that the temporal expansion of *upd* MA5 expression later into genital development may contribute to the origination of the posterior lobe (Sarah Smith, personal communication). Two final *trans* regulators of *Delta* identified in Chapter 4 included *Dorsocross2* (*Doc2*) and the *Doublesex*-regulated genital activity of *Drop* (*Dr*), but it is unclear whether they exhibit novel activities in posterior lobed species (Figure 6.1).

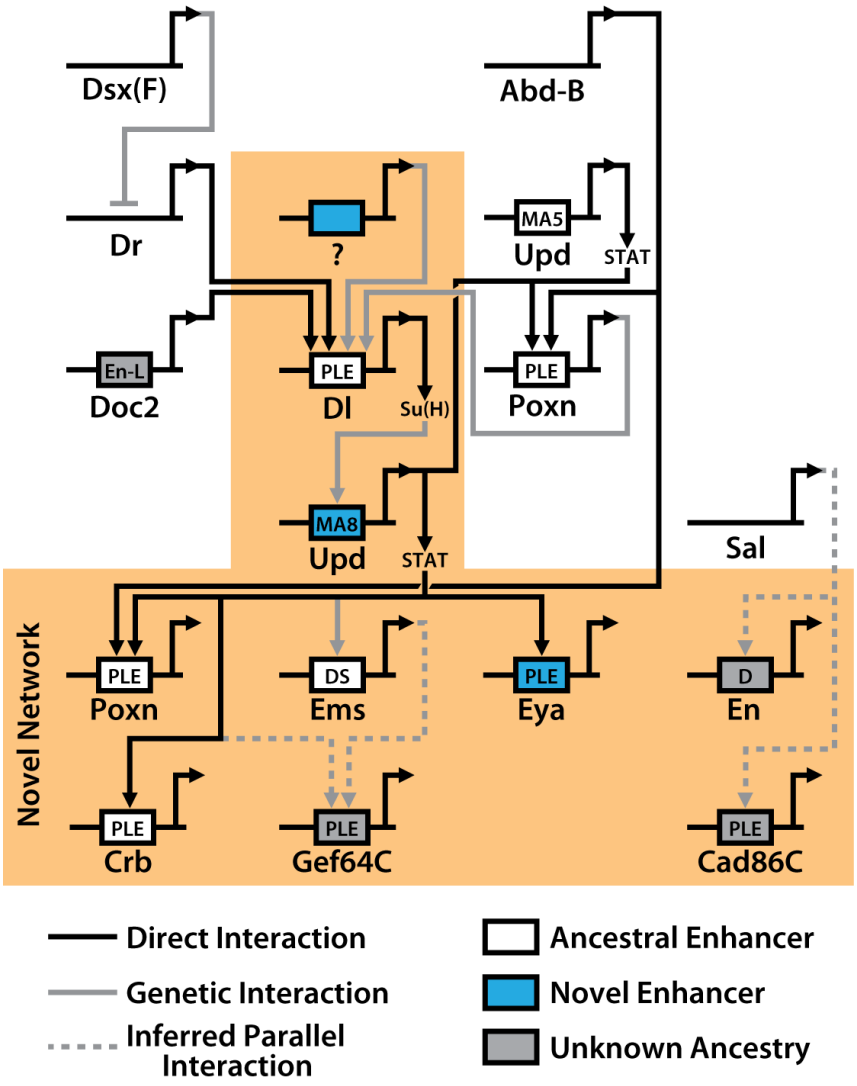

Figure 6.1. Diagram Illustrating the Posterior Lobe Gene Regulatory Network.

The orange area identifies components of the posterior lobe network that are uniquely deployed in posterior lobed species. Connections between genes that have evidence for direct binding of a transcription factor to an enhancer are drawn with a black line, connections with evidence for only an indirect genetic interaction are drawn with a gray line. Connections drawn with a dotted gray line refer to interactions that are inferred because the enhancer is shared with a network with known genetic interactions. Arrows denote an interaction that positively regulates expression while interactions ending with a line indicate inhibitory regulation. Enhancers that were ancestrally capable of driving expression in the posterior lobe are colored white. Enhancers that are uniquely active in posterior lobed species are colored blue. Enhancers for which the non-lobed enhancer has not been tested for

activity are colored gray. The question mark indicates a hypothesized novel *trans* regulator that interacts with *Poxn* to induce the expansion of *Delta*. Dsx(F): Female isoform of *Doublesex*.

6.1.2 Nodes of Several Unrelated Networks Contribute Regulatory Circuits to the Novel Posterior Lobe GRN

One of the most surprising discoveries to emerge from my investigation of the posterior lobe network has been the contribution of circuits from several completely separate and distinct tissues to the posterior lobe GRN. Enhancers and binding sites from morphologies as diverse as the posterior spiracle, the larval gut musculature, the wing hinge and the eye all have been recruited into the posterior lobe network (Figure 6.2). These findings may reveal an important fact about the evolution of unique and novel morphologies. Instead of creating a serial homologue of any one co-opted network's structure, the combination of multiple networks into the posterior lobe GRN may be responsible for the generation of its unique architecture.

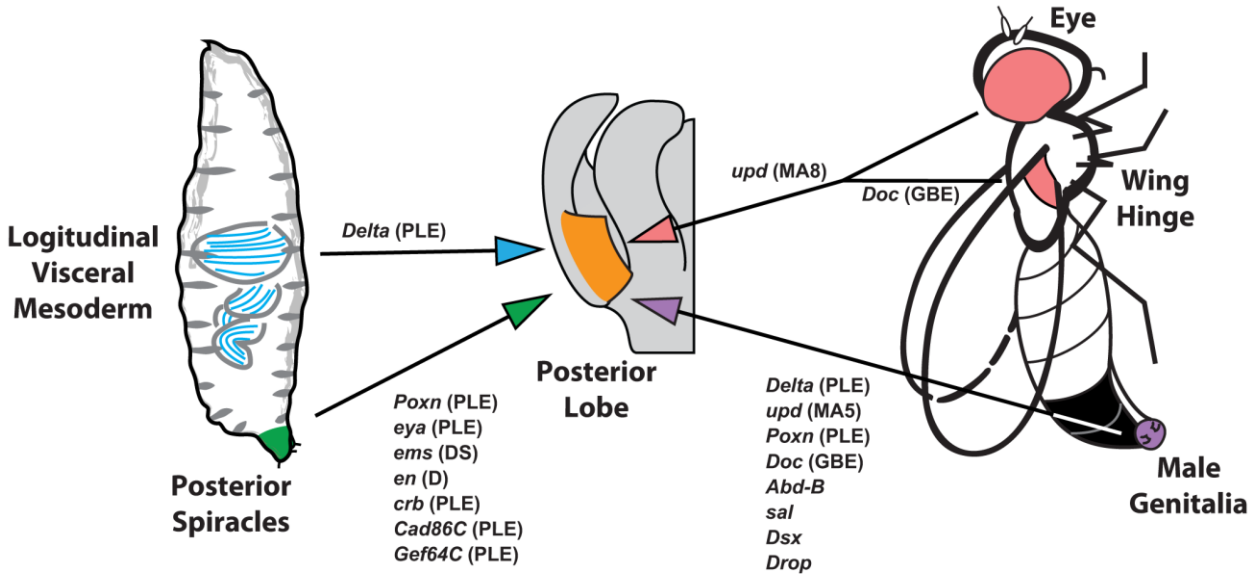

Figure 6.2. Model Depicting Ancestral Tissues from which Regulatory Circuitry was Co-opted During the Evolution of the Posterior Lobe.

Circuitry was co-opted into the posterior lobe network from several larval (left) and adult (right) morphologies, including the longitudinal visceral mesoderm (LVM) (blue), the posterior spiracle (green) the eye and the wing hinge (red) and the genitalia (orange). Genes and their enhancer active in the posterior lobe are listed beside their ancestral activity. (Black asterisk) Circuits that gained a new or expanded pattern of expression. (Orange asterisk) Circuits that required subsequent modification for deployment in the posterior lobe network.

6.1.3 Potential Targets for a Deeper Examination of Enhancer Co-option

One unexpected result from our analysis of co-opted transcription factor binding sites was that mutating individual binding sites completely ablated activity of the enhancer in the posterior lobe instead of merely reducing it (Figure 2.6). This stands in stark contrast to the Abd-B responsive abdominal enhancer of *bric-a-brac* (*bab*), in which a full 15 predicted Abd-B binding sites must be mutated to reduce activity to low levels (Williams et al., 2008). Why were we able to alter our

enhancer's activities so easily? It is possible that very ancient enhancers have evolved redundancy to maintain robust expression in the face of environmental or genetic perturbation, similar to the robustness thought to be provided by multiple enhancers that exhibit the same expression pattern (Frankel et al., 2010). Newly co-opted enhancers have not yet evolved the same level of redundancy for their novel activity and may therefore be more easily disrupted. To address this potential characteristic of co-opted enhancers, a more detailed analysis of the composition of co-opted posterior lobe enhancers is necessary. Two posterior lobe network genes that have been studied at a greater depth are the *crb* and *Poxn* posterior lobe enhancers.

Just prior to the publication of my posterior lobe network co-option study (Glassford et al., 2015), a deeper analysis of the *crb* spiracle enhancer was published (Pinto, Espinosa-Vazquez, Rivas, & Hombria, 2015). This study found that the Jak/STAT signaling inputs do not directly promote transcription, but instead indirectly promote expression by inhibiting the activity of a neighboring repressive region (Figure 6.3). Although we found that mutating these same STAT binding sites disrupts activity in the posterior lobe (Figure 2.6G-H), it is formally possible that the Jak/STAT binding sites may directly promote transcription in the posterior lobe, or that they repress an alternative region specific to the posterior lobe tissue. Studying the posterior lobe activity of the subfragments used in this new study would resolve these possibilities. Another interesting finding was that multiple Abd-B binding sites comprise the active subfragment (Pinto et al., 2015). Much like in the *bab* example, multiple Abd-B binding sites had to be mutated to disrupt the enhancer's activity. Further examination of the *crb* spiracle enhancer may provide a model to study the relative robustness of components regulating ancestral and novel activities.

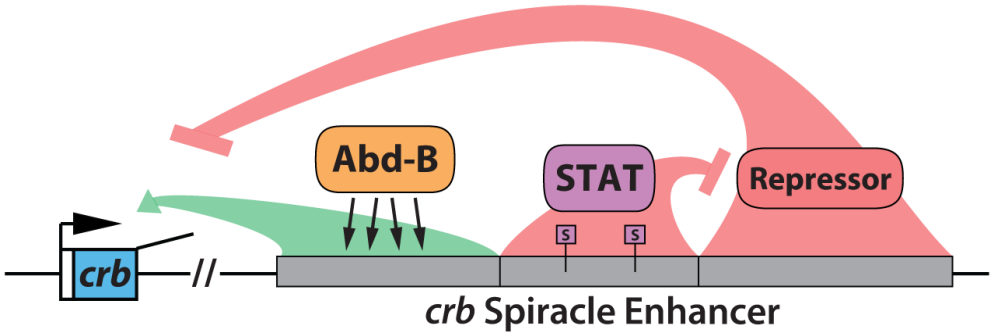

Figure 6.3. Jak/STAT Signaling Promotes *crb* Spiracle Enhancer Activity by Inhibiting a Repressive Subcomponent.

Schematic of the *crb* spiracle enhancer. Jak/STAT signaling integrates with the *crb* spiracle enhancer through two weak STAT binding sites (purple boxes) in a central subfragment of the enhancer. Binding of STAT (purple) releases a constitutively active subfragment proximal to the *crb* promoter (left) by repressing a more distal repressive subfagment (right). The constitutively active subfragment contains several Abd-B (orange) binding sites necessary for activity of the *crb* spiracle enhancer. Green arrows indicate positive regulatory interactions, red bars indicate repressive regulatory interactions. Figure is based on data presented in (Pinto et al., 2015).

The *Poxn* posterior lobe enhancer spans a relatively large span of DNA (~2.6kb), and an attempt to identify a smaller subfragment of this enhancer yielded reporters that drove ectopic expression in the genitalia and embryo (Figure 2.5A). This suggests that the *Poxn* posterior lobe enhancer is comprised of a core activating sequence (CAS) surrounded by repressive regulatory regions. An analysis of conserved sequences within the *Poxn* posterior lobe enhancer also supports this hypothesis. Alignment of the orthologous *Poxn* posterior lobe enhancer region from 15 *Drosophila* species identified 13 perfectly conserved blocks of sequence at least 10bp in length (Figure 5.A1). We scrambled several of these conserved regions by introducing a non-complementary transversion to every other base in the context of the enhancer. Several of these

mutations increased the activity of the reporter, suggesting that these sites encode repressive inputs to the enhancer. Interestingly, these mutations differentially altered posterior spiracle and posterior lobe activity, suggesting that the *Poxn* posterior lobe enhancer contains sites unique to its genital function (Figure 2.A1). Since these sites are deeply conserved throughout *Drosophila*, it is possible that the spiracle and lobe activities of *Poxn* were individualized since the origin of the posterior lobe. An analysis of these conserved blocks in a *Poxn* posterior lobe enhancer from a non-lobed species could confirm this hypothesis.

6.2 POTENTIAL UPSTREAM REGULATORS OF THE POSTERIOR LOBE NETWORK

Jak/STAT signaling (Glassford et al., 2015), and Notch Signaling (Chapter 4) are integral top-tier regulators of the posterior lobe gene regulatory network. It is possible that other intercellular signaling pathways are also important inputs to the posterior lobe network. The discovery of the lobe-associated pattern of *Delta* led me to revisit the role of several signaling ligands that had previously been discounted as posterior lobe genes such as the Wnt ligand *wingless* (wg) (See Addendum). Additionally, earlier work in our lab observed that flies mutant *Dachshous*, an upstream modulator of the Hippo signaling pathway (Harvey & Hariharan, 2012), altered the morphology of the posterior lobe (Rachel Pileggi, personal communication).

6.2.1 Approaches for Identifying Genes that Contribute to the Origination of the Posterior Lobe

Several techniques have been utilized in this dissertation to identify the upstream regulatory components of the posterior lobe GRN, including the candidate gene approach, and exploiting the rich background of knowledge concerning networks that may have been co-opted to the GRN. QTL mapping and introgression analysis have been used to identify genomic regions that contribute to the diversification of the posterior lobe (Macdonald & Goldstein, 1999; Masly, Dalton, Srivastava, Chen, & Arbeitman, 2011; Tanaka et al., 2015). Interestingly, a Genome Wide Association Study (GWAS) analysis of several QTLs contributing to the difference between the posterior lobes of *D. simulans* and *D. mauritiana* identified SNPs at several Wnt signaling components as candidate factors (Tanaka et al., 2015). QTL mapping and introgression analysis cannot be performed between lobed and non-lobed species as they cannot be crossed, but it is possible that the genes that contribute to the evolution of different shaped lobes may have had significant roles in its origination (Stern & Orgogozo, 2008). Hence, investigations of genes contributing to the lobe's divergence may unveil genes that were pivotal to this novel structure's origins.

6.3 DIVERSIFICATION OF THE DROSOPHILID EXTERNAL MALE GENITALIA

6.3.1 The Evolution of the Regulatory Locus of *Poxn*

The discovery that a modification of the *Poxn* posterior lobe enhancer alters both the clasper and posterior lobe morphologies resembles the phenomenon of "supergenes" that has been observed in butterfly mimicry rings. Traits controlled by supergenes depend on clusters of co-segregating and co-evolving genes that can exist in populations as stable polymorphisms (Schwander, Libbrecht, & Keller, 2014). Supergenes have been observed in complex phenotypes that combine behavioral and morphological alleles, allowing them co-segregate (Tuttle et al., 2016). As genital structures, the clasper (Acebes, Cobb, & Ferveur, 2003) and posterior lobe (Frazee & Masly, 2015) are integral components of *Drosophila* mating, and their modification is thought to contribute to the evolution of copulatory behavior; perhaps the pleiotropic modification of both the clasper and posterior lobe is a necessary component for a coherent behavioral modification. Supergenes typically exhibit reduced recombination to maintain the genetic link between separate alleles (Schwander et al., 2014), but a pleiotropic mutation such as the modification at *Poxn* would not require this mechanism. The *Poxn* locus contains other regulatory elements that govern separate aspects of *Drosophila* mating behavior, including copulatory length and male courtship (Boll & Noll, 2002) (Figure 6.4). Whether these other regions of *Poxn* were modified in the *D. simulans* or *D. mauritiana* lineages, and whether they are genetically linked to other modifications, warrants further study.

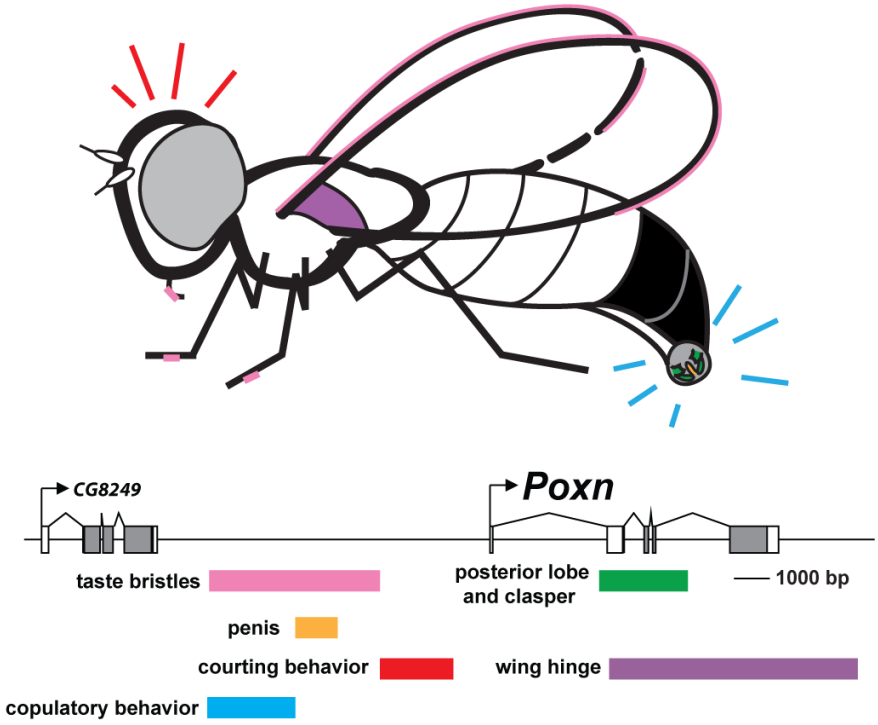

Figure 6.4. Map of Regulatory Regions that May Contribute to Male Courtship and Copulatory Behavior.

Selected regulatory regions of *Poxn* are displayed in a schematic of the *Poxn* regulatory locus (bottom) in colors that correspond to a model adult male fly (top). The wing hinge regulatory region is included as modification of this morphology may indirectly effect wing courtship display (Bennet-Clark, Ewing, & Bennet-Clark, 1968). Map of *Poxn* regulatory regions based on (Boll & Noll, 2002).

6.3.2 Conclusions

In this dissertation I have utilized an enhancer based approach to identify components of the novel posterior lobe network in an attempt to more deeply understand the evolution of novel morphologies. This methodology has allowed me to identify completely unexpected connections between gene regulatory networks and will likely be necessary to gain a complete understanding of the molecular mechanisms guiding the origination and diversification of novel morphologies.

BIBLIOGRAPHY

Abouheif, E. (1999). Establishing homology criteria for regulatory gene networks: prospects and challenges. *Novartis Foundation Symposium*, *222*, 207–21; discussion 222–5.

Acebes, A., Cobb, M., & Ferveur, J.-F. (2003). Species-specific effects of single sensillum ablation on mating position in Drosophila. *The Journal of Experimental Biology*, *206*(Pt 17), 3095–3100. http://doi.org/10.1242/jeb.00522

Ahrens, D., Schwarzer, J., & Vogler, A. P. (2014). The evolution of scarab beetles tracks the sequential rise of angiosperms and mammals. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1791), 20141470–20141470. http://doi.org/10.1098/rspb.2014.1470

Alberts, B., Johnson, A., Lewis, J., Morgan, D., Raff, M., Roberts, K., & Walter, P. (1989). *Molecular Biology of the Cell* (2nd ed.). Garland, New York.

Arnosti, D. N., & Kulkarni, M. M. (2005). Transcriptional enhancers: Intelligent enhanceosomes or flexible billboards? *Journal of Cellular Biochemistry*, *94*(5), 890–898. http://doi.org/10.1002/jcb.20352

Arrow, G. H. (1951). *Horned Beetles*. (W. D. Hinks, Ed.). The Hague, The Netherlands: W. Junk.

Bach, E. A., Ekas, L. A., Ayala-Camargo, A., Flaherty, M. S., Lee, H., Perrimon, N., & Baeg, G. H. (2007). GFP reporters detect the activation of the Drosophila JAK/STAT pathway in vivo. *Gene Expr Patterns*, *7*(3), 323–331. http://doi.org/S1567-133X(06)00156-6 [pii]10.1016/j.modgep.2006.08.003

Balhoff, J. P., & Wray, G. A. (2005). Evolutionary analysis of the well characterized endo16 promoter reveals substantial variation within functional sites. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(24), 8591–8596.

Barolo, S., Carver, L. a, & Posakony, J. W. (2000). GFP and beta-galactosidase transformation vectors for promoter/enhancer analysis in Drosophila. *BioTechniques*, *29*(4), 726, 728, 730, 732.

Barrio, R., Shea, M. J., Carulli, J., Lipkow, K., Gaul, U., Frommer, G., … Kafatos, F. C. (1996). The spalt-related gene of Drosophila melanogaster is a member of an ancient gene family,

defined by the adjacent, region-specific homeotic gene spalt. *Development Genes and Evolution*, *206*(5), 315–25. http://doi.org/10.1007/s004270050058

Beiman, M., Shilo, B.-Z., & Volk, T. (1996). Heartless , a Drosophila FGF receptor homolog , is essential for cell migration and establishment of several mesodermal lineages. *Genes & Development*, *10*, 2993–3002.

Bejsovec, A. (2013). Wingless/Wnt signaling in Drosophila: The pattern and the pathway. *Molecular Reproduction and Development*, *80*(11), 882–894. http://doi.org/10.1002/mrd.22228

Bennet-Clark, H. C. C., Ewing, A. W., & Bennet-Clark, H. C. C. (1968). The courtship songs of Drosophila. *Behaviour*, *31*(3), 288–301. http://doi.org/10.1163/156853968X00298

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

Blow, M. J., McCulley, D. J., Li, Z., Zhang, T., Akiyama, J. A., Holt, A., … Pennacchio, L. A. (2010). ChIP-Seq identification of weakly conserved heart enhancers. *Nature Genetics*, *42*(9), 806–810.

Bock, I. R., & Wheeler, M. (1972). The Drosophila melanogaster species group. *Univ. Texas Publ.*, *7213*, 1–102.

Boll, W., & Noll, M. (2002). The Drosophila Pox neuro gene: control of male courtship behavior and fertility as revealed by a complete dissection of all enhancers. *Development*, *129*(24), 5667–5681. http://doi.org/10.1242/dev.00157

Bonini, N. M., Bui, Q. T., Gray-Board, G. L., & Warrick, J. M. (1997). The Drosophila eyes absent gene directs ectopic eye formation in a pathway conserved between flies and vertebrates. *Development (Cambridge, England)*, *124*(23), 4819–26.

Bonner, J. T. (1998). The origins of multicellularity. *Integrative Biology: Issues, News, and Reviews*, *1*(1), 27–36. http://doi.org/10.1002/(SICI)1520-6602(1998)1:1<27::AID-INBI4>3.3.CO;2-Y

Bradley, R. K., Li, X.-Y., Trapnell, C., Davidson, S., Pachter, L., Chu, H. C., … Eisen, M. B. (2010). Binding Site Turnover Produces Pervasive Quantitative Changes in Transcription Factor Binding between Closely Related Drosophila Species. *PLoS Biology*, *8*(3), 16.

Bridgham, J. T., Carroll, S. B., & Thornton, J. W. (2006). Evolution of hormone-receptor complexity by molecular exploitation. *Science*, *312*(5770), 97–101. http://doi.org/312/5770/97 [pii]10.1126/science.1123348

Britten, R. J. (1996). Cases of ancient mobile element DNA insertions that now affect gene regulation. *Molecular Phylogenetics and Evolution*, *5*(1), 13–17.

http://doi.org/10.1006/mpev.1996.0003

Broihier, H. T., Moore, L. A., Doren, M. Van, Newman, S., & Lehmann, R. (1998). zfh-1 is required for germ cell migration and gonadal mesoderm development in Drosophila, *666*, 655–666.

Brown, S., & Castelli Gair Hombría, J. (2000). Drosophila grain encodes a GATA transcription factor required for cell rearrangement during morphogenesis. *Development (Cambridge, England)*, *127*(22), 4867–4876.

Burtis, K. C., & Baker, B. S. (1989). Drosophila doublesex gene controls somatic sexual differentiation by producing alternatively spliced mRNAs encoding related sex-specific polypeptides. *Cell*, *56*(6), 997–1010. http://doi.org/10.1016/0092-8674(89)90633-8

Cande, J. D., Chopra, V. S., & Levine, M. S. (2009). Evolving enhancer-promoter interactions within the tinman complex of the flour beetle, Tribolium castaneum. *Development*, *136*(18), 3153–3160. http://doi.org/10.1242/dev.038034

Carroll, S. B. (2008). Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell*, *134*(1), 25–36. http://doi.org/10.1016/j.cell.2008.06.030

Carroll, S. B., Grenier, J. K., & Weatherbee, S. D. (2004). *From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design* (2nd ed.). Wiley-Blackwell.

Celniker, S. E., Keelan, D. J., & Lewis, E. B. (1989). The molecular genetics of the bithorax complex of Drosophila: characterization of the products of the Abdominal-B domain. *Genes & Development*, *3*(9), 1424–1436.

Chan, Y. F., Marks, M. E., Jones, F. C., Villarreal Jr., G., Shapiro, M. D., Brady, S. D., … Kingsley, D. M. (2010). Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer. *Science*, *327*(5963), 302–305. http://doi.org/science.1182213 [pii]10.1126/science.1182213

Chatterjee, S. S., Uppendahl, L. D., Chowdhury, M. a, Ip, P.-L., & Siegal, M. L. (2011). The female-specific doublesex isoform regulates pleiotropic transcription factors to pattern genital development in Drosophila. *Development (Cambridge, England)*, *138*(6), 1099–1109. http://doi.org/10.1242/dev.055731

Cheng, Y., Brunner, A. L., Kremer, S., DeVido, S. K., Stefaniuk, C. M., & Kassis, J. A. (2014). Co-regulation of invected and engrailed by a complex array of regulatory sequences in Drosophila. *Developmental Biology*, *395*(1), 131–43. http://doi.org/10.1016/j.ydbio.2014.08.021

Chung, H., Bogwitz, M. R., McCart, C., Andrianopoulos, A., Ffrench-Constant, R. H., Batterham, P., & Daborn, P. J. (2007). Cis-regulatory elements in the accord retrotransposon result in tissue-specific expression of the Drosophila melanogaster insecticide resistance gene Cyp6g1. *Genetics*, *175*(3), 1071–1077. http://doi.org/10.1534/genetics.106.066597

Chuong, E. B., Elde, N. C., & Feschotte, C. (2016). Regulatory evolution of innate immunity through co-option of endogenous retroviruses. *Science*, *351*(6277), 1083–1087. http://doi.org/10.1126/science.aad5497

Coyne, J. (1993). The Genetics of an Isolating Mechanism between 2 Sibling Species of Drosophila. *Evolution; International Journal of Organic Evolution*, *47*(3), 778–788. http://doi.org/10.2307/2410183

Coyne, J. A., Rux, J., & David, J. R. (1991). Genetics of morphological differences and hybrid sterility between Drosophila sechellia and its relatives. *Genetical Research*, *57*(2), 113–122.

Cretekos, C. J., Wang, Y., Green, E. D., Martin, J. F., Rasweiler, J. J. th, & Behringer, R. R. (2008). Regulatory divergence modifies limb length between mammals. *Genes Dev*, *22*(2), 141–151. http://doi.org/22/2/141 [pii]10.1101/gad.1620408

Crick, F. H. (1968). The origin of the genetic code. *Journal of Molecular Biology*, *38*(3), 367–79.

Daborn, P., Boundy, S., Yen, J., Pittendrigh, B., & Ffrench-Constant, R. H. (2001). DDT resistance in Drosophila correlates with Cyp6g1 over-expression and confers cross-resistance to the neonicotinoid imidacloprid. *Molecular Genetics and Genomics*, *266*(4), 556–563. http://doi.org/10.1007/s004380100531

Dalton, D., Chadwick, R., & McGinnis, W. (1989). Expression and embryonic function of empty spiracles: a Drosophila homeo box gene with two patterning functions on the anterior-posterior axis of the embryo. *Genes & Development*, *3*(12A), 1940–56.

Dambly-Chaudiere, C., Jamet, E., Burri, M., Bopp, D., Basler, K., Hafen, E., … Noll, M. (1992). The paired box gene pox neuro: a determinant of poly-innervated sense organs in Drosophila. *Cell*, *69*(1), 159–172. http://doi.org/0092-8674(92)90127-X [pii]

Davidson, E. H. (2001). *Genomic regulatory systems : development and evolution*. San Diego: Academic Press.

Davidson, E. H. (2006). The regulatory genome: gene regulatory networks in development and evolution. *Developmental Biology*. http://doi.org/10.1016/j.ydbio.2007.08.009

de Souza, F. S. J., Franchini, L. F., & Rubinstein, M. (2013). Exaptation of transposable elements into novel cis-regulatory elements: is the evidence always strong? *Molecular Biology and Evolution*, *30*(6), 1239–51. http://doi.org/10.1093/molbev/mst045

Dermitzakis, E. T., & Clark, A. G. (2002). Evolution of transcription factor binding sites in Mammalian gene regulatory regions: conservation and turnover. *Molecular Biology and Evolution*, *19*(7), 1114–1121.

Dong, P. D., Todi, S. V, Eberl, D. F., & Boekhoff-Falk, G. (2003). Drosophila spalt/spalt-related mutants exhibit Townes-Brocks' syndrome phenotypes. *Proc Natl Acad Sci U S A*, *100*(18), 10293–10298. http://doi.org/10.1073/pnas.1836391100 1836391100 [pii]

Duboule, D. (2007). The rise and fall of Hox gene clusters. *Development (Cambridge, England)*, *134*(14), 2549–2560. http://doi.org/10.1242/dev.001065

Dynan, W. S. (1989). Modularity in promoters and enhancers. *Cell*, *58*(1), 1–4. http://doi.org/10.1016/0092-8674(89)90393-0

Eberhard, W. G. (1985). *Sexual selection and animal genitalia*. Cambridge, Mass.: Harvard University Press.

Eichenlaub, M. P., & Ettwiller, L. (2011). De Novo Genesis of Enhancers in Vertebrates, *9*(11). http://doi.org/10.1371/journal.pbio.1001188

Ekker, S. C., Jackson, D. G., Von Kessler, D. P., Sun, B. I., Young, K. E., & Beachy, P. A. (1994). The degree of variation in DNA sequence recognition among four Drosophila homeotic proteins. *The The European Molecular Biology Organization Journal*, *13*(15), 3551–3560.

Emlen, D. J., Lavine, L. C., & Ewen-Campen, B. (2007). On the origin and evolutionary diversification of beetle horns. *Proceedings of the National Academy of Sciences of the United States of America*, *104*, 8661–8668. http://doi.org/10.1073/pnas.0701209104

Emlen, D. J., Szafran, Q., Corley, L. S., & Dworkin, I. (2006). Insulin signaling and limb-patterning: candidate pathways for the origin and evolutionary diversification of beetle "horns." *Heredity*, *97*(3), 179–191. http://doi.org/10.1038/sj.hdy.6800868

Emlen, D. J., Warren, I. a., Johns, a., Dworkin, I., & Lavine, L. C. (2012). A Mechanism of Extreme Growth and Reliable Signaling in Sexually Selected Ornaments and Weapons. *Science*, *337*(6096), 860–864. http://doi.org/10.1126/science.1224286

Erwin, D. H. (2015). Novelty and innovation in the history of life. *Current Biology*, *25*(19), R930–R940. http://doi.org/10.1016/j.cub.2015.08.019

Erwin, D. H., & Davidson, E. H. (2009). The evolution of hierarchical gene regulatory networks. *Nature Reviews. Genetics*, *10*(2), 141–148. http://doi.org/10.1038/nrg2499

Estrada, B., & Sánchez-Herrero, E. (2001). The Hox gene Abdominal-B antagonizes appendage development in the genital disc of Drosophila. *Development Cambridge England*, *128*(3), 331–339.

Ezkurdia, I., Juan, D., Rodriguez, J. M., Frankish, A., Diekhans, M., Harrow, J., … Tress, M. L. (2014). Multiple evidence strands suggest that theremay be as few as 19 000 human protein-coding genes. *Human Molecular Genetics*, *23*(22), 5866–5878. http://doi.org/10.1093/hmg/ddu309

Foronda, D., Estrada, B., de Navas, L., & Sanchez-Herrero, E. (2006). Requirement of Abdominal-A and Abdominal-B in the developing genitalia of Drosophila breaks the posterior downregulation rule. *Development*, *133*(1), 117–127. http://doi.org/dev.02173 [pii]10.1242/dev.02173

Frankel, N., Davis, G. K., Vargas, D., Wang, S., Payre, F., & Stern, D. L. (2010). Phenotypic robustness conferred by apparently redundant transcriptional enhancers. *Nature*, *466*(7305), 490–3. http://doi.org/10.1038/nature09158

Frankel, N., Erezyilmaz, D. F., McGregor, A. P., Wang, S., Payre, F., & Stern, D. L. (2011). Morphological evolution caused by many subtle-effect substitutions in regulatory DNA. *Nature*, *474*(7353), 598–603. http://doi.org/10.1038/nature10200

Frazee, S. R., & Masly, J. P. (2015). Multiple sexual selection pressures drive the rapid evolution of complex morphology in a male secondary genital structure. *Ecology and Evolution*, *5*(19), 4437–4450. http://doi.org/10.1002/ece3.1721

Frazzetta, T. H. (2012). Flatfishes, Turtles, and Bolyerine Snakes: Evolution by Small Steps or Large, or Both? *Evolutionary Biology*, *39*(1), 30–60. http://doi.org/10.1007/s11692-011-9142-7

Gao, F., & Davidson, E. H. (2008). Transfer of a large gene regulatory apparatus to a new developmental address in echinoid evolution. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(16), 6091–6. http://doi.org/10.1073/pnas.0801201105

Gaszner, M., & Felsenfeld, G. (2006). Insulators: exploiting transcriptional and epigenetic mechanisms. *Nature Reviews. Genetics*, *7*(9), 703–713. http://doi.org/10.1038/nrg1925

Geffeney, S. L., Fujimoto, E., Brodie, E. D., & Ruben, P. C. (2005). Evolutionary diversification of TTX-resistant sodium channels in a predator-prey interaction. *Nature*, *434*(7034), 759–763.

Glassford, W. J., Johnson, W. C., Dall, N. R., Smith, S. J., Liu, Y., Boll, W., … Rebeiz, M. (2015). Co-option of an Ancestral Hox-Regulated Network Underlies a Recently Evolved Morphological Novelty. *Developmental Cell*, *34*(5), 520–531. http://doi.org/10.1016/j.devcel.2015.08.005

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

Groth, A. C., Fish, M., Nusse, R., & Calos, M. P. (2004). Construction of transgenic Drosophila by using the site-specific integrase from phage phiC31. *Genetics*, *166*(4), 1775–82.

Guruharsha, K. G., Kankel, M. W., & Artavanis-tsakonas, S. (2012). The Notch signalling system : recent insights into the complexity of a conserved pathway. *Nature Publishing Group*, *13*(9), 654–666. http://doi.org/10.1038/nrg3272

Hackerl, U., Kaufmann, E., Hartmann, C., Jurgens, G., Knochel, W., & Jacklel, H. (1995). The Drosophila fork head domain protein crocodile is required for the establishment of head structures, *14*(21), 5306–5317.

Halder, G., Callaerts, P., & Gehring, W. J. (1995). Induction of ectopic eyes by targeted expression of the eyeless gene in Drosophila. *Science (New York, N.Y.)*, *267*(5205), 1788–1792. http://doi.org/10.1126/science.7892602

Hardison, R. C. (2000). Conserved noncoding sequences are reliable guides to regulatory elements. *Trends in Genetics*, *16*(9), 369–372.

Harrison, D. A., McCoon, P. E., Binari, R., Gilman, M., & Perrimon, N. (1998). Drosophila unpaired encodes a secreted protein that activates the JAK signaling pathway. *Genes & Development*, *12*(20), 3252–3263.

Harvey, K. F., & Hariharan, I. K. (2012). The Hippo pathway. *Cold Spring Harbor Perspectives in Biology*, *4*(8). http://doi.org/10.1101/cshperspect.a011288

Higashijima, S., Michiue, T., Emori, Y., & Saigo, K. (1992). Subtype determination of Drosophila embryonic external sensory organs by redundant homeo box genes BarH1 and BarH2. *Genes & Development*, *6*(6), 1005–1018. http://doi.org/10.1101/gad.6.6.1005

Hong, J.-W., Hendrix, D. A., & Levine, M. S. (2008). Shadow Enhancers as a Source of Evolutionary Novelty. *Science*, *321*(5894), 1314. http://doi.org/10.1126/science.1160631

Hu, N., & Castelli Gair Hombría, J. (1999). Study of the posterior spiracles of Drosophila as a model to understand the genetic and cellular mechanisms controlling morphogenesis. *Developmental Biology*, *214*(1), 197–210. http://doi.org/10.1006/dbio.1999.9391

Hyman, L. H. (1955). *The Invertebrates: Echinodermata*. New York: McGraw-Hill.

Ismat, A., Schaub, C., Reim, I., Kirchner, K., Schultheis, D., & Frasch, M. (2010). HLH54F is required for the specification and migration of longitudinal gut muscle founders from the caudal mesoderm of Drosophila. *Development (Cambridge, England)*, *137*(18), 3107–3117. http://doi.org/10.1242/dev.046573

Jagadeeshan, S., & Singh, R. S. (2006). A time-sequence functional analysis of mating behaviour and genital coupling in Drosophila: role of cryptic female choice and male sex-drive in the evolution of male genitalia. *Journal of Evolutionary Biology*, *19*(4), 1058–70. http://doi.org/10.1111/j.1420-9101.2006.01099.x

Jenett, A., Rubin, G. M., Ngo, T.-T. B., Shepherd, D., Murphy, C., Dionne, H., … Zugates, C. T. (2012). A GAL4-driver line resource for Drosophila neurobiology. *Cell Reports*, *2*(4), 991–1001. http://doi.org/10.1016/j.celrep.2012.09.011

Johnson, J. E., Birren, S. J., & Anderson, D. J. (1990). Two rat homologues of Drosophila achaete-scute specifically expressed in neuronal precursors. *Nature*, *346*(6287), 858–861. http://doi.org/10.1038/346858a0

Jones, B., & McGinnis, W. (1993). The regulation of empty spiracles by Abdominal-B mediates an abdominal segment identity function. *Genes & Development*, *7*(2), 229–240. http://doi.org/10.1101/gad.7.2.229

Jürgens, G., Wieschaus, E., Nüsslein-Volhard, C., & Kluding, H. (1984). Mutations affecting the pattern of the larval cuticle in Drosophila melanogaster II. Zygotic loci on the third chromosome. *Developmental Biology*, *193*, 283–295.

Kadonaga, J. T. (2004). Regulation of RNA Polymerase II Transcription by Sequence-Specific DNA Binding Factors. *Cell*, *116*(2), 247–257. http://doi.org/10.1016/S0092-8674(03)01078-X

Kamimura, Y. (2010). Copulation anatomy of Drosophila melanogaster (Diptera: Drosophilidae): wound-making organs and their possible roles. *Zoomorphology*, *129*(3), 163–174. http://doi.org/10.1007/s00435-010-0109-5

Kelsh, R., Dawson, I., & Akam, M. (1993). An analysis of abdominal-B expression in the locust Schistocerca gregaria. *Development (Cambridge, England)*, *117*(1), 293–305.

Keys, D. N., Lewis, D. L., Selegue, J. E., Pearson, B. J., Goodrich, L. V, Johnson, R. L., … Carroll, S. B. (1999). Recruitment of a hedgehog Regulatory Circuit in Butterfly Eyespot Evolution. *Science*, *283*(5401), 532–534. http://doi.org/10.1126/science.283.5401.532

Kijimoto, T., Moczek, A. P., & Andrews, J. (2012). Diversification of doublesex function underlies morph-, sex-, and species-specific development of beetle horns. *Proceedings of the National Academy of Sciences*, *109*(50), 20526–20531. http://doi.org/10.1073/pnas.1118589109

Kopp, A., & True, J. R. (2002). Evolution of male sexual characters in the oriental Drosophila melanogaster species group. *Evolution & Development*, *4*(4), 278–91. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12168620

Koshikawa, S., Giorgianni, M. W., Vaccaro, K., Kassner, V. a, Yoder, J. H., Werner, T., & Carroll, S. B. (2015). Gain of cis-regulatory activities underlies novel domains of wingless gene expression in Drosophila. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(24), 7524–9. http://doi.org/10.1073/pnas.1509022112

Kühnlein, R. P., Frommer, G., Friedrich, M., Gonzalez-Gaitan, M., Weber, A., Wagner-Bernholz, J. F., … Schuh, R. (1994). spalt encodes an evolutionarily conserved zinc finger protein of novel structure which provides homeotic gene function in the head and tail region of the Drosophila embryo. *The EMBO Journal*, *13*(1), 168–79.

Kuraku, S., Usuda, R., & Kuratani, S. (2005). Comprehensive survey of carapacial ridge-specific genes in turtle implies co-option of some regulatory genes in carapace evolution. *Evol Dev*, *7*(1), 3–17. http://doi.org/EDE05002 [pii]10.1111/j.1525-142X.2005.05002.x

Laurie, C. C., True, J. R., Liu, J., & Mercer, J. M. (1997). An Introgression Analysis of Quantitative Trait Loci That Contribute to a Morphological Difference Between, (Figure 1).

Lee, H., Lee, H., Zaffran, S., & Frasch, M. (2005). Visceral Musculature, (April 2016), 0–17. http://doi.org/10.1007/0-387-32963-3

Levine, M. S. (2010). Transcriptional enhancers in animal development and evolution. *Current Biology*, *20*(17), R754–R763. http://doi.org/10.1016/j.cub.2010.06.070

Liu, F., & Posakony, J. W. (2012). Role of architecture in the function and specificity of two Notch-regulated transcriptional enhancer modules. *PLoS Genetics*, *8*(7), e1002796. http://doi.org/10.1371/journal.pgen.1002796

Loots, G. G., Locksley, R. M., Blankespoor, C. M., Wang, Z. E., Miller, W., Rubin, E. M., & Frazer, K. A. (2000). Identification of a coordinate regulator of interleukins 4, 13, and 5 by cross-species sequence comparisons. *Science*, *288*(5463), 136–140.

Lovegrove, B., Simões, S., Rivas, M. L., Sotillos, S., Johnson, K., Knust, E., … Hombría, J. C.-G. (2006). Coordinated control of cell adhesion, polarity, and cytoskeleton underlies Hox-induced organogenesis in Drosophila. *Current Biology : CB*, *16*(22), 2206–16. http://doi.org/10.1016/j.cub.2006.09.029

Ludwig, M. Z., Bergman, C., Patel, N. H., & Kreitman, M. (2000). Evidence for stabilizing selection in a eukaryotic enhancer element. *Nature*, *403*(6769), 564–567. http://doi.org/10.1038/35000615

Ludwig, M. Z., Kittler, R., White, K. P., & Kreitman, M. (2011). Consequences of Eukaryotic Enhancer Architecture for Gene Expression Dynamics, Development, and Fitness. *PLoS Genetics*, *7*(11), e1002364. http://doi.org/10.1371/journal.pgen.1002364

Ludwig, M. Z., Patel, N. H., & Kreitman, M. (1998). Functional analysis of eve stripe 2 enhancer evolution in Drosophila: rules governing conservation and change. *Development*, *125*(5), 949–958. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=9449677\nhttp://dev.biologists.org/content/125/5/949.full.pdf

Lupianez, D. G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., … Mundlos, S. (2015). Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell*, *161*(5), 1012–1025. http://doi.org/10.1016/j.cell.2015.04.004

Lynch, V. J., Leclerc, R. D., May, G., & Wagner, G. P. (2011). Transposon-mediated rewiring of gene regulatory networks contributed to the evolution of pregnancy in mammals. *Nature Genetics*, *43*(11), 1154–9. http://doi.org/10.1038/ng.917

Macdonald, S. J., & Goldstein, D. B. (1999). A quantitative genetic analysis of male sexual traits distinguishing the sibling species Drosophila simulans and D. sechellia. *Genetics*, *153*(4), 1683–99. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1460840&tool=pmcentrez&rendertype=abstract

Maier, D. (2006). Hairless: The ignored antagonist of the Notch signalling pathway. *Hereditas*, *143*(2006), 212–221. http://doi.org/10.1111/j.2007.0018-0661.01971.x

Masly, J. P., Dalton, J. E., Srivastava, S., Chen, L., & Arbeitman, M. N. (2011). The genetic

basis of rapidly evolving male genital morphology in Drosophila. *Genetics*, *189*(1), 357–374. http://doi.org/10.1534/genetics.111.130815

McCauley, B. S., Wright, E. P., Exner, C., Kitazawa, C., & Hinman, V. F. (2012). Development of an embryonic skeletogenic mesenchyme lineage in a sea cucumber reveals the trajectory of change for the evolution of novel structures in echinoderms. *EvoDevo*, *3*(1), 17. http://doi.org/10.1186/2041-9139-3-17

McGregor, A. P., Orgogozo, V., Delon, I., Zanet, J., Srinivasan, D. G., Payre, F., & Stern, D. L. (2007). Morphological evolution through multiple cis-regulatory mutations at a single gene. *Nature*, *448*(7153), 587–590. http://doi.org/nature05988 [pii]10.1038/nature05988

Moczek, A. P., & Nagy, L. M. (2005). Diverse developmental mechanisms contribute to different levels of diversity in horned beetles. *Evolution & Development*, *7*(3), 175–85. http://doi.org/10.1111/j.1525-142X.2005.05020.x

Moczek, A. P., & Rose, D. J. (2009). Differential recruitment of limb patterning genes during development and diversification of beetle horns. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(22), 8992–7. http://doi.org/10.1073/pnas.0809668106

Moczek, A. P., Rose, D. J., Sewell, W., & Kesselring, B. R. (2006). Conservation, innovation, and the evolution of horned beetle diversity. *Dev Genes Evol*, *216*(11), 655–665. http://doi.org/10.1007/s00427-006-0087-2

Monteiro, A., Glaser, G., Stockslager, S., Glansdorp, N., & Ramos, D. (2006). Comparative insights into questions of lepidopteran wing pattern homology. *BMC Developmental Biology*, *6*, 52. http://doi.org/10.1186/1471-213X-6-52

Monteiro, A., & Podlaha, O. (2009). Wings, horns, and butterfly eyespots: how do complex traits evolve? *PLoS Biology*, *7*(2), e37. http://doi.org/10.1371/journal.pbio.1000037

Muller, G. B., & Wagner, G. P. (1991). Novelty in Evolution: Restructuring the Concept. *Annual Review of Ecology and Systematics*.

Muñoz Descalzo, S., & Martinez Arias, A. (2012). The structure of Wntch signalling and the resolution of transition states in development. *Seminars in Cell and Developmental Biology*, *23*(4), 443–449. http://doi.org/10.1016/j.semcdb.2012.01.012

Nachman, M. W., Hoekstra, H. E., & D'Agostino, S. L. (2003). The genetic basis of adaptive melanism in pocket mice. *Proc Natl Acad Sci U S A*, *100*(9), 5268–5273. http://doi.org/10.1073/pnas.0431157100 [pii]

Neitz, M., Neitz, J., & Jacobs, G. H. (1991). Spectral tuning of pigments underlying red-green color vision. *Science*, *252*(5008), 971–974.

Ni, X., Zhang, Y. E., Nègre, N., Chen, S., Long, M., & White, K. P. (2012). Adaptive Evolution and the Birth of CTCF Binding Sites in the Drosophila Genome. *PLoS Biology*, *10*(11).

http://doi.org/10.1371/journal.pbio.1001420

Nichols, S. a, Dirks, W., Pearse, J. S., & King, N. (2006). Early evolution of animal cell signaling and adhesion genes. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(33), 12451–12456. http://doi.org/10.1073/pnas.0604065103

Nitta, K. R., Jolma, A., Yin, Y., Morgunova, E., Kivioja, T., Akhtar, J., … Taipale, J. (2015). Conservation of transcription factor binding specificities across 600 million years of bilateria evolution. *eLife*, *2015*(4), 1–20. http://doi.org/10.7554/eLife.04837

Noyes, M. B., Christensen, R. G., Wakabayashi, A., Stormo, G. D., Brodsky, M. H., & Wolfe, S. A. (2008). Analysis of Homeodomain Specificities Allows the Family-wide Prediction of Preferred Recognition Sites. *Cell*, *133*(7), 1277–1289. http://doi.org/10.1016/j.cell.2008.05.023

Oliver, J. C., Tong, X.-L., Gall, L. F., Piel, W. H., & Monteiro, A. (2012). A single origin for nymphalid butterfly eyespots followed by widespread loss of associated gene expression. *PLoS Genetics*, *8*(8), e1002893. http://doi.org/10.1371/journal.pgen.1002893

Ong, C.-T., & Corces, V. G. (2014). CTCF: an architectural protein bridging genome topology and function. *Nature Reviews. Genetics*, *15*(4), 234–46. http://doi.org/10.1038/nrg3663

Orgel, L. E. (1968). Evolution of the genetic apparatus. *Journal of Molecular Biology*, *38*(3), 381–93.

Orr, H. A. (2005). The genetic theory of adaptation: a brief history. *Nature Reviews Genetics*, *6*(2), 119–127.

Ortlund, E. A., Bridgham, J. T., Redinbo, M. R., & Thornton, J. W. (2007). Crystal structure of an ancient protein: evolution by conformational epistasis. *Science*, *317*(5844), 1544–1548. http://doi.org/1142819 [pii] 10.1126/science.1142819

Patel, N. H., Martin-Blanco, E., Coleman, K. G., Poole, S. J., Ellis, M. C., Kornberg, T. B., & Goodman, C. S. (1989). Expression of engrailed proteins in arthropods, annelids, and chordates. *Cell*, *58*(5), 955–68.

Perrimon, N., Pitsouli, C., & Shilo, B.-Z. (2012). Signaling mechanisms controlling cell fate and embryonic patterning. *Cold Spring Harbor Perspectives in Biology*, *4*(8), a005975. http://doi.org/10.1101/cshperspect.a005975

Perry, M. W., Boettiger, A. N., Bothma, J. P., & Levine, M. S. (2010). Shadow enhancers foster robustness of Drosophila gastrulation. *Current Biology*, *20*(17), 1562–1567.

Peterson, B. K., Hare, E. E., Iyer, V. N., Storage, S., Conner, L., Papaj, D. R., … Eisen, M. B. (2009). Big Genomes Facilitate the Comparative Identification of Regulatory Elements. *PLoS ONE*, *4*(3), 12.

Pfeiffer, B. D., Jenett, A., Hammonds, A. S., Ngo, T.-T. B., Misra, S., Murphy, C., … Rubin, G.

M. (2008). Tools for neuroanatomy and neurogenetics in Drosophila. *Proc Natl Acad Sci U S A*, *105*(28), 9715–9720. http://doi.org/0803697105 [pii]10.1073/pnas.0803697105

Pinto, P. B., Espinosa-Vazquez, J. M., Rivas, M. L., & Hombria, J. C. (2015). JAK/STAT and Hox Dynamic Interactions in an Organogenetic Gene Cascade. *PLoS Genet*, *11*(7), e1005412. http://doi.org/10.1371/journal.pgen.1005412

Prelich, G. (2012). Gene Overexpression: Uses, Mechanisms, and Interpretation. *Genetics*, *190*(3), 841–854. http://doi.org/10.1534/genetics.111.136911

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

Prudic, K. L., Stoehr, A. M., Wasik, B. R., & Monteiro, A. (2015). Eyespots deflect predator attack increasing fitness and promoting the evolution of phenotypic plasticity. *Proceedings of the Royal Society of London B: Biological Sciences*, *282*(1798), 20141531. http://doi.org/10.1098/rspb.2014.1531

Rebeiz, M., Castro, B., Liu, F., Yue, F., & Posakony, J. W. (2012). Ancestral and conserved cis-regulatory architectures in developmental control genes. *Developmental Biology*, *362*(2), 282–294. http://doi.org/10.1016/j.ydbio.2011.12.011

Rebeiz, M., Jikomes, N., Kassner, V. A., & Carroll, S. B. (2011). Evolutionary origin of a novel gene expression pattern through co-option of the latent activities of existing regulatory sequences. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(25), 10036–43. http://doi.org/10.1073/pnas.1105937108

Rebeiz, M., Patel, N. H., & Hinman, V. F. (2015). Unraveling the Tangled Skein : The Evolution of Transcriptional Regulatory Networks in Development. http://doi.org/10.1146/annurev-genom-091212-153423

Rebeiz, M., Pool, J. E., Kassner, V. A., Aquadro, C. F., & Carroll, S. B. (2009). Stepwise modification of a modular enhancer underlies adaptation in a Drosophila population. *Science (New York, N.Y.)*, *326*(5960), 1663–7. http://doi.org/10.1126/science.1178357

Rebeiz, M., & Posakony, J. W. (2004). GenePalette: a universal software tool for genome sequence visualization and analysis. *Dev Biol*, *271*(2), 431–438. http://doi.org/10.1016/j.ydbio.2004.04.011S0012160604002805 [pii]

Reim, I., Lee, H., & Frasch, M. (2003). The T-box-encoding Dorsocross genes function in amnioserosa development and the patterning of the dorsolateral germ band downstream of Dpp, 3187–3204. http://doi.org/10.1242/dev.00548

Reinhardt, K. (2009). Natural selection and genital variation: A role for the environment, parasites and sperm ageing? *Genetica*, *138*(1), 119–127. http://doi.org/10.1007/s10709-009-9360-4

Robertson, H. M. (1988). Mating Asymmetries and Phylogeny in the Drosophila melanogaster. *Pacific Science*, *42*, 1–2.

Rockman, M. V, & Wray, G. A. (2002). Abundant raw material for cis-regulatory evolution in humans. *Mol Biol Evol*, *19*(11), 1991–2004.

Sagan, L. (1967). On the origin of mitosing cells. *Journal of Theoretical Biology*, *14*(3), 255–74.

Schwander, T., Libbrecht, R., & Keller, L. (2014). Supergenes and complex phenotypes. *Current Biology*, *24*(7), R288–R294. http://doi.org/10.1016/j.cub.2014.01.056

Shastry, B. S. (1995). Overexpression of genes in health and sickness. A bird's eye view. *Comparative Biochemistry and Physiology Part B Biochemistry Molecular Biology*, *112*(1), 1–13.

Simpson, T. L. (1984). *The Cell Biology of Sponges*. New York: Springer US.

Small, S., Blair, A., & Levine, M. S. (1992). Regulation of even-skipped stripe 2 in the Drosophila embryo. *The EMBO Journal*, *11*(11), 4047–4057.

Snell-Rood, E. C., Cash, A., Han, M. V., Kijimoto, T., Andrews, J., & Moczek, A. P. (2011). Developmental decoupling of alternative phenotypes: Insights from the transcriptomes of horn-polyphenic beetles. *Evolution*, *65*(1), 231–245. http://doi.org/10.1111/j.1558-5646.2010.01106.x

Srivastava, M., Simakov, O., Chapman, J., Fahey, B., Gauthier, M. E. a, Mitros, T., … Rokhsar, D. S. (2010). The Amphimedon queenslandica genome and the evolution of animal complexity. *Nature*, *466*(7307), 720–6. http://doi.org/10.1038/nature09201

Stearns, F. W. (2010). One hundred years of pleiotropy: A retrospective. *Genetics*, *186*(3), 767–773. http://doi.org/10.1534/genetics.110.122549

Stern, D. L. (2000). Evolutionary developmental biology and the problem of variation. *Evolution*, *54*(4), 1079–1091.

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

Stevens, M. (2005). The role of eyespots as anti-predator mechanisms, principally demonstrated in the Lepidoptera. *Biological Reviews*, *80*(4), 573–588. http://doi.org/10.1017/S1464793105006810

Stork, N. E., McBroom, J., Gely, C., & Hamilton, A. J. (2015). New approaches narrow global species estimates for beetles, insects, and terrestrial arthropods. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(24), 7519–23. http://doi.org/10.1073/pnas.1502408112

Swami, M. (2010). Transcription: Shadow enhancers confer robustness. *Nature Reviews.*

Genetics, *11*(7), 454. http://doi.org/10.1038/nrg2818

Swanson, C. I., Evans, N. C., & Barolo, S. (2010). Structural rules and complex regulatory circuitry constrain expression of a Notch- and EGFR-regulated eye enhancer. *Developmental Cell*, *18*(3), 359–370. http://doi.org/10.1016/j.devcel.2009.12.026

Swanson, C. I., Schwimmer, D. B., & Barolo, S. (2011). Rapid evolutionary rewiring of a structurally constrained eye enhancer. *Current Biology*, *21*(14), 1186–1196. http://doi.org/10.1016/j.cub.2011.05.056

Symmons, O., & Spitz, F. (2013). From remote enhancers to gene regulation: charting the genome's regulatory landscapes. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *368*(1620), 20120358. http://doi.org/10.1098/rstb.2012.0358

Tamura, K., Subramanian, S., & Kumar, S. (2004). Temporal patterns of fruit fly (Drosophila) evolution revealed by mutation clocks. *Molecular Biology and Evolution*, *21*(1), 36–44. http://doi.org/10.1093/molbev/msg236

Tanaka, K. M., Hopfen, C., Herbert, M. R., Schlotterer, C., Stern, D. L., Masly, J. P., … Nunes, M. D. S. (2015). Genetic Architecture and Functional Characterization of Genes Underlying the Rapid Diversification of Male External Genitalia Between Drosophila simulans and Drosophila mauritiana. *Genetics*, *200*(1), 357–369. http://doi.org/10.1534/genetics.114.174045

Tepass, U., & Knust, E. (1993). Crumbs and stardust act in a genetic pathway that controls the organization of epithelia in Drosophila melanogaster. *Developmental Biology*, *159*(1), 311–26. http://doi.org/10.1006/dbio.1993.1243

Thanos, D., & Maniatis, T. (1995). Virus induction of human IFN beta gene expression requires the assembly of an enhanceosome. *Cell*, *83*(7), 1091–100.

Tomancak, P., Beaton, A., Weiszmann, R., Kwan, E., Shu, S., Lewis, S. E., … Rubin, G. M. (2002). Systematic determination of patterns of gene expression during Drosophila embryogenesis, 1–14.

Tuttle, E. M., Bergland, A. O., Korody, M. L., Brewer, M. S., Newhouse, D. J., Minx, P., … Balakrishnan, C. N. (2016). Divergence and Functional Degradation of a Sex Chromosome-like Supergene. *Current Biology*, *26*(3), 344–350. http://doi.org/10.1016/j.cub.2015.11.069

Valentine, J. W. (2003). *Keywords and Concepts in Evolutionary Developmental Biology*. (B. K. Hall & W. M. Olson, Eds.) (Hall, B. K). Cambridge: Harvard University Press.

Vietri Rudan, M., Barrington, C., Henderson, S., Ernst, C., Odom, D. T., Tanay, A., & Hadjur, S. (2015). Comparative Hi-C Reveals that CTCF Underlies Evolution of Chromosomal Domain Architecture. *Cell Reports*, *10*(8), 1297–1309. http://doi.org/10.1016/j.celrep.2015.02.004

Wagner, G. P., & Zhang, J. (2011). Fundamental concepts in genetics: The pleiotropic structure of the genotype–phenotype map: the evolvability of complex organisms. *Nature Publishing Group*, *12*(3), 204–213. http://doi.org/10.1038/nrg2949

Weinreich, D. M., Delaney, N. F., Depristo, M. A., & Hartl, D. L. (2006). Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science*, *312*(5770), 111–114. http://doi.org/312/5770/111 [pii]10.1126/science.1123539

Weinreich, D. M., Watson, R. A., & Chao, L. (2005). Perspective: Sign epistasis and genetic constraint on evolutionary trajectories. *Evolution*, *59*(6), 1165–1174.

Werner, T., Koshikawa, S., Williams, T. M., & Carroll, S. B. (2010). Generation of a novel wing colour pattern by the Wingless morphogen. *Nature*, *464*(7292), 1143–8. http://doi.org/10.1038/nature08896

Williams, T. M., Selegue, J. E., Werner, T., Gompel, N., Kopp, A., & Carroll, S. B. (2008). The regulation and evolution of a genetic switch controlling sexually dimorphic traits in Drosophila. *Cell*, *134*(4), 610–623. http://doi.org/S0092-8674(08)00840-4 [pii]10.1016/j.cell.2008.06.052

Wittkopp, P. J. (2005). Genomic sources of regulatory variation in cis and in trans. *Cellular and Molecular Life Sciences : CMLS*, *62*(16), 1779–83. http://doi.org/10.1007/s00018-005-5064-9

Woolfe, A., Goodson, M., Goode, D. K., Snell, P., McEwen, G. K., Vavouri, T., … Elgar, G. (2005). Highly Conserved Non-Coding Sequences Are Associated with Vertebrate Development. *PLoS Biology*, *3*(1), e7.

Wray, G. A. (2003). Transcriptional regulation and the evolution of development. *The International Journal of Developmental Biology*, *47*, 675–684.

Wray, G. A. (2007). The evolutionary significance of cis-regulatory mutations. *Nature Reviews. Genetics*, *8*(3), 206–16. http://doi.org/10.1038/nrg2063

Wray, G. A., Hahn, M. W., Abouheif, E., Balhoff, J. P., Pizer, M., Rockman, M. V, & Romano, L. A. (2003). The evolution of transcriptional regulation in eukaryotes. *Mol Biol Evol*, *20*(9), 1377–1419. http://doi.org/10.1093/molbev/msg140msg140 [pii]

Yan, R., Small, S., Desplan, C., Dearolf, C. R., & Darnell, J. E. (1996). Identification of a Stat Gene That Functions in Drosophila Development. *Cell*, *84*(3), 421–430. http://doi.org/10.1016/S0092-8674(00)81287-8

Yassin, A., & Orgogozo, V. (2013). Coevolution between Male and Female Genitalia in the Drosophila melanogaster Species Subgroup. *PloS One*, *8*(2), e57158. http://doi.org/10.1371/journal.pone.0057158

Yoder, J. H., & Carroll, S. B. (2006). The evolution of abdominal reduction and the recent origin of distinct Abdominal-B transcript classes in Diptera. *Evolution & Development*, *8*(3), 241–

51. http://doi.org/10.1111/j.1525-142X.2006.00095.x

Zaffran, S., Küchler, A., Lee, H., & Frasch, M. (2001). biniou (FoxF), a central component in a regulatory network controlling visceral mesoderm development and midgut morphogenesis in Drosophila, 2900–2915. http://doi.org/10.1101/gad.917101.tebrate

Zeng, Z., Liu, J., Stam, L. F., Kao, C., Mercer, J. M., & Laurie, C. C. (2000). Two Drosophila Species, *310*(January), 299–310.

Zuckerkandl, E. (1976). Programs of Gene Action and Progressive Evolution. In M. Goodman, R. Tashian, & J. Tashian (Eds.), *Molecular Anthropology SE - 20* (Vol. 62, pp. 387–447). Springer US. http://doi.org/10.1007/978-1-4615-8783-5_20