RAPID EVOLUTION OF *CIS*-REGULATORY ARCHITECTURE AND ACTIVITY IN THE DROSOPHILA *YELLOW* GENE

by

Gizem Kalay

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Molecular, Cellular and Developmental Biology)
in the University of Michigan
2012

Doctoral Committee:

Associate Professor Patricia J. Wittkopp, Chair
Professor Kenneth M. Cadigan
Professor Robert J. Denver
Professor Jianzhi Zhang
Associate Professor Scott E. Barolo

“*Yurtta Baris, Dunyada Baris*”

“*Peace at home, peace in the world*”

Mustafa Kemal Ataturk

Dedications

To my two grandmothers who thought me the value of working hard and
loving life regardless of any condition.

To my father from whom I learnt the value of wisdom, patience and tolerance.

To my mother for her endless love and teaching me how to be practical.

To my sister and brother, whom I know never stopped believing in me and will
continue to do so no matter what happens.

To all my dearest friends without whose support I would not be where I am
personally and professionally.

To Jonathan, for being the sweetest person I have ever met and making my life
such a warm and happy place.

Acknowledgments

I would like to, first of all, thank Patricia Wittkopp for being such a good teacher and sharing her great mind and open heart with me, always making me feel very lucky to be working with her. Initially, being able to benefit from her mentorship firsthand and later on from the highly intellectual and friendly lab environment she created were invaluable in my scientific and personal growth throughout graduate school. I also would like to thank Xiaowei Heng and Mackenzie Dome for their contributions to Chapter 2 and 3, respectively, of this thesis and for being such good colleagues, collaborators, and friends; Soochin Cho for supporting me personally and professionally during first two tough years of graduate school; Barry WIlliams for his generous share of knowledge on yeast as I juggled through my first project; my committee members Ken Cadigan, Bob Denver, Jianzhi Zhang, and Scott Barolo, for their invaluable support and guidance in the past five years; Bart Deplancke and his lab members for their contribution to Chapter 3 and their friendship; all past and current members of the Wittkopp Lab, particularly, Belinda Haerum, Adam Neidert, Jonathan Gruber, Arielle Cooley and Elizabeth Walker for making the lab such a fun and happy place to work for me; Karishma Collette, Anita Guedea and Pia Bagamasbad for being there for me physically and emotionally; Duygu Tosun and Ozlem Keles to whom I know I can and will always rely on; Juanita Cabello and Sally Haines for ever so gently showing and teaching me how to lead a happy life and for being very generous with their warmth and love; Atilgan Yilmaz, Nezaket Turkel, Ilker Oztop and Didem Goz, and my friendship with them for being so sincere, easy, tolerant and FUN; Mina and Banu Riolo for loving me for what I am; Burcu Bayirli and Chris Riolo for their encouragement and support; Kara Vogel and her boys for being the most open-hearted family I have ever seen and caring for me so sincerely; Seda Guler and Berra Sagun, whom I am so glad are a part of my life; Ita Reyes for helping me discover my love for yoga and being one of most wonderful people I have ever met; Ali Benli for cheering me up and sharing his positive energy towards life each time I talked to him; Emily, Chris and Curtis Morgan, Andrea Morris and Collin Spencer for being such sweet and supportive friends. The research in this thesis was funded by University of Michigan, National Science Foundation, Rackham Graduate School and Ecole Polytechnique Federale de Lausanne.

Table of Contents

List of Figures

List of Tables

Abstract

In the last 10 years, an increasing number of case studies showed that changes in *cis*-regulatory elements, mainly enhancers, are one of the main causes of altered phenotypes, but the mechanisms underlying enhancer evolution remain to be elucidated. More specifically, what is the relationship between changes in enhancer sequence, transcription factor binding and activity? In this thesis, I used evolution of *yellow* enhancers among Drosophila species as a model to shed light onto how *cis*-regulatory architecture and activity change over time. I first identified the enhancer activities lying in the 5' intergenic and intronic regions of *yellow* from six Drosophila species spanning the Drosophila evolutionary history, using a reporter gene assay. I found that *yellow* epidermal-cell and wing-vein, but not bristle enhancers, have different positions, with respect to the coding sequence, in different Drosophila species. This was the first systematic demonstration of altered enhancer position between species and suggested that enhancer position can be labile. Sequence comparisons failed to show any indication of translocation or duplication suggesting gradual compensatory changes in the transcription factor binding profiles of *yellow* enhancers is the likely mechanism underlying altered enhancer position. Subsequent subdivision of *yellow* 5' intergenic and intronic regions showed a complex distribution of enhancer activities among sub-elements, where some drove expression in patterns that were not part of the expression pattern driven by the full region. Existence of such "cryptic" epidermal-cell enhancer activities suggests that *yellow cis*-regulatory regions were primed for facilitating the rapid evolutionary changes in the position and activities of this enhancer. Lastly, for the first time, I identified a large set of candidate transcription factors binding to *yellow* enhancers. This thesis shows that position and activity of *yellow* enhancers diverged rapidly among species, perhaps by taking advantage of the cryptic activities lying in the *yellow cis*-regulatory sequences. Further *in vitro* and *in vivo* tests validating the direct binding of the identified transcription factors on *yellow* enhancers and characterizing their functional effects on *yellow* expression among species can elucidate the evolutionary changes underlying altered position and activity of the particular *yellow* enhancers.

Chapter 1

Introduction

Genes need to be expressed in order to give rise to the product they encode. This expression step is highly regulated at multiple levels, including but not limited to change in chromatin packaging, transcription, post-transcriptional processing, translation and post-translational processing. Each level of control adds a layer of complexity to gene regulation which allows fine tuning of gene expression. It is known that changes in gene expression may give rise to changes in phenotype and even to disease states. These changes in gene expression can occur due to changes at any of the steps controlling it. Hence, it is crucial to understand the intricacies of the regulatory mechanism underlying gene expression in order to understand organismal evolution and physiology.

Controllers of gene expression can be categorized as *cis* and *trans* acting factors. At the transcriptional level, *cis*-acting factors mainly consist of *cis*-regulatory sequences, such as promoters, enhancers and insulators. These are, typically, non-coding DNA regions that harbor a certain combination of short sequence motifs, which are distributed with a particular composition within the *cis*-regulatory element and are bound by *trans* acting factors that affect transcription (i.e., transcription factors). Among the *cis*-regulatory elements, enhancers harbor the majority of the information that determine when, where and how much the corresponding gene will be transcribed. As a result, as compared to other *cis*-regulatory elements, enhancers affect gene expression more prominently. Transcription factors, on the other hand, are diffusible molecules, typically proteins, that affect transcription by binding to the short sequence motifs, i.e., binding sites, found in the *cis*-regulatory elements in a sequence-specific manner. Overall, specific interactions between *cis*-regulatory elements and transcription factors control a gene's transcription, hence, changes in the activities of either or both can alter expression patterns.

Between the changes in the activities of enhancers versus transcription factors, it is proposed that the former is likely to cause phenotypic changes more often than the latter one. This is because enhancers have a modular organization where each module controls a gene's transcription in a specific developmental time and tissue/cell type. Hence, for instance, a change in the activity of an enhancer module would be very specific since it would affect the expression of only the corresponding gene, only in a subset of the cells and only in the specific developmental time frame the gene is expressed. On the other hand, a change in the activity of a transcription factor would be broader and affect the expression of all the genes it controls. In other words, as compared to the changes in transcription factors, changes in enhancers are less pleiotropic. As a result, changes in enhancer sequences are less likely to be destructive to the organism and, hence, are less likely to be selected against by natural selection. Because of this it is proposed that changes in the activity of enhancer sequences, as compared to changes in the activity of *trans* acting factors, make up the majority of the genetic changes underlying morphological evolution.

In the following sections I will first give an overview of the molecular mechanism enhancers work through in order to regulate transcription. This will be followed by how enhancer sequence and activity changes over time, and what some possible consequences of these changes are at the organismal level. Afterwards, different ways enhancers can be gained and lost will be discussed with examples. Next will come how different enhancer modules, in other words the whole *cis*-

regulatory architecture of a gene, can be organized and altered over time. And lastly, I will talk about the function as well as the *cis* and *trans* acting factors that affect regulation of the Drosophila *yellow* gene and the knowledge accumulated so far about its *cis*-regulatory architecture and its evolution.

How do enhancers interact with different *cis*-regulatory elements and *trans* factors to affect gene transcription in eukaryotes?

In enhancer sequences, the information for a particular expression pattern is encoded in the form of transcription factor binding sites, where the type and number of as well as the spacing between these binding sites constitute the grammar of the particular enhancer. The size of an enhancer on average ranges between 100 base pairs (bp) to several kilo base pairs (kbp). It can be located in the 5' or 3' intergenic sequence of a gene as well as in intronic regions or hundreds of kilo base pairs away in the genome (Kleinjan and van Heyningen 2005), even in different chromosomes (Lomvardas et al. 2006).

For proper functioning of enhancers, their appropriate interaction with transcription factors as well as other *cis*-regulatory elements is crucial. Hence, before describing the mechanism of how enhancers regulate transcription, it is important to introduce the other components of the eukaryotic *cis*-regulatory system such as promoters and boundary elements (insulators). This will be followed by a summary of the current model for the mechanism of enhancer function. This summary will include, but will not be limited to, how the interactions among transcription factors and cofactors and their effects on chromatin states and nucleosome organization influence enhancers and, ultimately, transcription.

Promoters consist of a core promoter and a promoter-proximal region. Core promoter can extend up to 40 base pairs(bp) in either direction of the +1 position (Blackwood and Kadonaga 1998; Baumann et al. 2010). It contains transcription initiation site as well as several of other sequence motifs like the TATA box, initiator (Inr), TFIIB recognition element (BRE), downstream core promoter element (DPE) and the downstream core element (DCE). Different combinations of these motifs are found in core promoters, where they are essential for proper transcriptional activation. For instance, although not exclusively, DPE and Inr, together, are commonly found in TATA-less promoters and are shown to be functional counterparts to TATA-box, such that they are bound by the general transcription factor TFIID and are required for transcriptional activation (Smale and Kadonaga 2003; Baumann et al. 2010). Under appropriate conditions, the core promoter is sufficient to initiate transcription by recruiting the RNA polymerase II (Pol II) basal transcriptional machinery, which consists of RNA polymerase II and various general transcription factors. These general transcription factors are responsible for proper positioning of the polymerase as well as interacting with other specific transcription factors and cofactors.

Proximal promoter region, located immediately 5' of the core promoter (approximately from -40 to -250 relative to the transcription start site), is essential for the proper expression of genes (Blackwood and Kadonaga 1998; Baumann et al. 2010). It contains binding sites for tissue specific transcription factors and can be encompassing enhancer elements. Even though the promoter proximal region is tissue specific, due to its close proximity of the to the core promoter, the two elements tend to be collectively referred to as the “promoter”.

Insulators are another type of *cis*-regulatory element that when located between a promoter and an enhancer, block the gene activating or repressing effects of the enhancer; however if they are flanking the enhancer-promoter pair they do not affect the interaction between the two. Different insulator elements vary in sequence as well as the specific proteins that bind them (Nègre et al. 2010). In fact the blocking activity of insulators is dependent on their binding by sequence-specific proteins (e.g. CCCTC-binding factor in vertebrates (CTCF), Supressor of Hairy wing (Su(Hw)) in Drosophila) (Bell et al. 1999; Parnell et al. 2003). Insulators can also block spreading of the silencing effects of heterochromatin and some, but not all, insulator elements are capable of executing both functions (West et al. 2002).

As *cis*-regulatory elements that can affect transcription from distant locations in the genome, with respect to their cognate gene, enhancers require a particular mechanism to be able to directly interact with the core promoter. The most commonly accepted model for the mechanism of how enhancers interact with the core promoter is called “looping” (Blackwood and Kadonaga 1998; Schoenfelder et al. 2010). In this model, the DNA between an enhancer and the promoter loops out in order to facilitate the direct interaction between the *trans* factors that are bound to the enhancer element and those that are bound to the promoter. It is not known through what mechanism DNA looping occurs, but one prominent model is called “facilitated tracking”, where an enhancer bound complex comprising transcription factors and cofactors moves along DNA via small steps, as opposed to a continuous movement, while scanning the DNA for insulator and promoter elements (Blackwood and Kadonaga 1998). A 30-subunit protein complex called the Mediator, which is crucial for activation of transcription in many genes by mediating the interaction between the Pol II basal transcriptional machinery and gene specific transcription factors (Malik and Roeder 2010), cohesin and a particular type of non-coding RNA molecule (eRNA) have been suggested to take part in the looping of the DNA (Ong and Corces 2011). Regardless of the exact mechanism of how the DNA region containing the enhancer loops out onto the promoter, colocalization of many enhancer and promoter regions has been shown by chromosome conformation capture (3C) assays (Dekker et al. 2002; Visel et al. 2009). This assay involves crosslinking to capture interacting loci, DNA fragmentation, intermolecular ligation and PCR analysis of the resulting ligated products and measures the frequency of two genomic loci detected in close proximity to each other. This frequency is inversely correlated to the distance between the two genomic loci. Hence, for instance, if a distal enhancer is detected in close proximity to the promoter more frequently than expected given its distance from the promoter, then this suggests direct interaction between this distal enhancer and the promoter.

Another “challenge” caused by enhancers not always being located in close proximity to their cognate gene, and at times nearby other genes that they do not regulate, is potential unspecific effects of the enhancers on genes other than their cognate gene. There are two known mechanisms that prevent unspecific enhancer-gene interactions. One of them is enhancer-promoter specificity. That is to say, some enhancers are selective in what type of promoter they interact with. For instance in the *Drosophila melanogaster* genome, autoregulatory element-1 (AE1) enhancer is equidistant from both *Sex combs reduced* (*Scr*) and *fushi tarazu* (*ftz*) promoters, but it selectively activates *ftz* expression only (Ohtsuki et al. 1998). In vitro studies showed that in the absence of a competing TATA-containing promoter, AE1 is capable of activating a TATA-less promoter, but in the presence of both a TATA-containing and TATA-less promoters, AE1 preferentially activates the TATA-containing promoter. Intriguingly, *ftz* promoter harbors a TATA box whereas *Scr* promoter does not. This shows the importance of enhancer-promoter specificity for proper regulation of gene expression. Not all enhancers, however, show selectivity on what type of promoter they interact with. For instance, Butler and Kadonaga (Butler and Kadonaga 2001) compared the activities of eighteen *D. melanogaster* enhancers when they interacted with a TATA versus a DPE containing core promoter connected to a common reporter gene, where each time the compared transgenes were inserted in the same genomic location. They found that only four of the eighteen enhancers interacted with one and not the other type of promoter, whereas the remaining fourteen did not show a noticeable preference between the two types of promoters to activate transcription. It is possible that the enhancers that do not show promoter specificity, prevent unspecific enhancer-gene interactions by using the second mechanism, where an insulator element between an enhancer and a promoter blocks the enhancer from acting on the promoter.

Besides their interactions with other *cis*-regulatory elements, in order to function properly, it is also crucial for enhancers to interact with the appropriate combination of *trans* factors. These *trans* factors include transcription factors, which bind to enhancer DNA directly as well as cofactors that do not bind to DNA themselves, but interact with DNA binding transcription factors (Näär et al. 2001; Orphanides and Reinberg 2002).

At least some of the transcription factors and cofactors recruited to an enhancer have enzymatic activities that facilitate chromatin remodeling (Blackwood and Kadonaga 1998; Näär et al. 2001; Orphanides and Reinberg 2002). This is important for decondensation of chromatin to allow assembly of the basal transcriptional machinery at the promoter. The *trans* factors with enzymatic activity facilitate chromosome remodeling by catalyzing ATP dependent nucleosome repositioning to make the enhancer and promoter regions accessible to transcription factors. The enzymatic activities of the *trans* factors can also covalently modify various histone proteins (e.g., acetylation or deacetylation of H3 and H4) and lead to, for instance, change in the electric charge of the histone proteins which affects how tightly the negatively charged DNA is wrapped around histones. This in turn alters the accessibility of transcription factor binding sites to their binding factors and influences expression.

Besides chromosome remodeling, some transcription factors and cofactors have enzymatic activities and can catalyze phosphorylation and acetylation of other transcription factors, modifying proteins' activities and properties (Blackwood and Kadonaga 1998; Näär et al. 2001; Orphanides and Reinberg 2002). For instance, *in vitro*, the form of p53 tumor suppressor protein that is virtually inactive in site-specific DNA binding, gets activated only after being acetylated by its coactivator (p300) (Gu and Roeder 1997). These types of modifications on proteins that are recruited to the enhancer elements are crucial for proper expression of genes.

Studies have also proposed that enhancers may be targeting their corresponding genes to specialized nuclear domains that have high local concentrations of transcription factors. For instance, in mammalian erythroid cells, the chromosomal territory harboring the ß-globin genes is relocalized in the nucleus via an extrusion. This extrusion occurs prior to high-level ß-globin gene expression and was found to be dependent on the existence of the enhancers of this locus (Locus Control Region - LCR) that are also required for high-level ß-globin gene expression (Ragoczy et al. 2003; Bulger and Groudine 2011). This suggested that LCR directs the ß-globin locus to a part of the nucleus that perhaps has necessary transcription factors for the activation of ß-globin transcription.

Enhancers work together with other *cis*-regulatory elements as well as *trans* acting factors in order to regulate transcription in a time, place and quantity specific manner. Some of the underlying mechanisms of enhancer function, and how enhancers interact with other *cis*-regulatory elements and *trans* acting factors have been revealed. However, this area of research still bears important questions that will be elucidated in the coming years.

How do sequence and function of enhancers change over time?

It has been shown that changes in enhancer sequences can alter gene expression, which in turn can lead to changes in phenotypes (reviewed in (Carroll 2008; Wittkopp and Kalay 2012)) or even disease states (reviewed in (Visel et al. 2009)). Hence for a complete understanding of organismal evolution and physiology, it is important to understand how enhancer sequence and activity change over time. More specifically, what types of mutations alter enhancer activity?

It is a well-known fact that continual occurrence of mutations subject DNA sequences to incessant change. Some of these sequence changes occur in functional DNA regions and can alter their activity. The change in the activity of the functional element may be beneficial or disruptive for the organism, making the causal mutation beneficial or disruptive. Other mutations occur either in putatively nonfunctional DNA regions or in functional regions but without leading to a change in activity. These mutations are regarded to be neutral. Over time beneficial mutations are selected for, or kept in the population, whereas disruptive mutations are selected against, removed from the population. Neutral mutations, on the other hand, are not selected for or against and are likely to accumulate in the population over time.

In coding DNA sequences, mutations that do not change the amino acid sequence (synonymous) of a gene are generally more likely to be neutral than mutations that change the amino-acid sequence (non-synonymous) of the gene. There are exceptions, however; for instance, a synonymous mutation can change the codon encoding for the same amino acid to one that is not preferred by the organism and this may change the expression level of the corresponding protein, which in turn may be beneficial or disruptive for the organism (Hershberg and Petrov 2008; Sauna and Kimchi-Sarfaty 2011).

As functional elements in the genome, mutations in enhancer sequences can also be beneficial, deleterious or neutral for the activity of the particular enhancer, and potentially for the whole organism. Compared to the coding sequences, however, it is harder to make an assessment on whether any given mutation is beneficial, deleterious or neutral just by looking at the enhancer sequences. This is because there is no known universal code underlying enhancer activity. That is to say, for any given enhancer sequence one cannot distinguish essential pieces from nonessential ones in the absence of functional assays. There are several different reasons lying behind this:

Firstly, transcription factor binding sites are known to be the essential subunits of enhancers, but the knowledge on these sites is still limited, i.e., there is a substantial number of unidentified transcription factor binding sites (Bulyk 2003).

Secondly, transcription factor binding sites are degenerate, meaning one or more nucleotides in them are interchangable with other nucleotides (Stormo 2000).

Thirdly, there may be other essential sequences in enhancer elements, besides transcription factor binding sites. Some of these can affect nucleosome positioning as exemplified by the work of Tirosh and colleagues (Tirosh et al. 2008). The authors showed that changes in the sequences flanking an essential transcription factor binding site (Ste1) in the promoters of several genes from three yeast species (*Saccharomyces cerevisiae*, *S. paradoxus* and *S. mikatae*) affected nucleosome occupancy of this transcription factor binding site, which strongly correlated with the changes in the expression level of the genes investigated. There are also other sequences in enhancer regions, that have not been shown to bind to transcription factors or nucleosomes so far, but affect enhancer activity, suggesting they may have functions that are not yet identified (Swanson et al. 2010).

Last but not the least, the distance between transcription factor binding sites as well as their orientation with respect to each other and the coding sequence, may be essential for proper enhancer activity (Senger et al. 2004; Williams et al. 2008), however there is currently no universal rule about the orientation and composition of transcription factors in enhancer sequences and their relations to enhancer activity. Hence, once again, investigating the enhancer sequence alone cannot give a conclusive assessment without conducting functional assays.

As compared to the coding sequences, the above features of enhancers make it hard to identify the functional units that came together to form a particular enhancer. Additionally, the same features make it hard to find enhancers in the genome. Fortunately, availability of genomic sequences from multiple species proved very useful for both of the above-mentioned problems.

One method that makes use of genomic sequences from multiple species and has proved very helpful in investigating enhancers is called phylogenetic footprinting (Prabhakar et al. 2006; Elgar and Vavouri 2008; Loots 2008; Woolfe and Elgar 2008). In this method, researchers compare sequences from multiple species to find highly similar regions in the genome. Since disruptive mutations in functional elements are selected against over time, parts of the genome that have highly similar sequences are thought to have function. Among the regions with highly similar sequences between species, the non-coding ones are typically regarded as candidate enhancers ready to be tested for activity. The majority of the time these conserved non-coding sequences show tissue specific enhancer activity *in vivo* (Shin et al. 2005; Woolfe et al. 2005; Pennacchio et al. 2006; Peterson et al. 2009; O'Quin et al. 2011). Hence, phylogenetic footprinting is a useful method for finding enhancers in the genome.

In addition to finding enhancers in the genome, sequence similarity data is useful also for conferring functional similarity. This is because many orthologous enhancers that have higher sequence similarity to each other than to other parts of the genome also have conserved activity (Hadzhiev et al. 2007; Peterson et al. 2009).

This is not always the case, however. That is to say, sequences of two functionally homologous enhancers (located in physically homologous positions and harbor comparable activities affecting the same gene) can be highly similar, but the few changes between them can be sufficient to alter their activity (Goode et al. 2011). When this is the case, changes in enhancer activity can be mapped to a handful of nucleotide differences using sequence comparisons between closely related species as well as an outgroup species (Booth et al. 2010; Frankel et al. 2011; Rebeiz et al. 2011a). This outgroup species should be evolutionarily equidistant from the two other species being compared and it should also have enhancer activity similar to one of them. This way one can determine the divergent nucleotides in the enhancer sequence that drives the divergent expression pattern. In order to find the actual causal mutations these candidate causal sequence changes then can be tested by being introduced into the enhancer of the opposite species individually, and in groups, and testing the activity of the hybrid enhancer using a reporter gene in transgenic animals.

For instance, despite the high sequence similarity between the *shaven baby* E6 enhancer from *D. melanogaster* and *D. sechelia*, the *D. sechelia* E6 element fails to drive dorsolateral expression in stage 14 embryos, which in turn leads to lack of larval trichomes in the corresponding body part in *D. sechelia*. This is due only to a handful (14) of sequence substitutions observed between *D. melanogaster* and *D.secehelia* E6 elements. Replacing the *D.sechelia*-like causative mutations in the E6 element with their *D. melanogaster* counterparts is sufficient to restore dorsolateral expression of reporter gene in Drosophila embryos, and consequently to recover larval dorsolateral trichomes (Frankel et al. 2011).

Once causal mutations are identified, one can investigate the mechanism through which these mutations led to change in enhancer activity, e.g.; whether they lead to gain or loss of transcription factor binding sites (Gompel et al. 2005a; Jeong et al. 2006a), or perhaps change in nucleosome positioning (Tirosh et al. 2008; Tsankov et al. 2010), etc. Multiple causal mutations can also be tested individually as well as mutually to test whether there is epistasis between them or their effects on expression are additive (Frankel et al. 2011).

Sequences of functionally homologous enhancers, however, do not always have higher similarity than the sequence around them. In fact, there are many cases where sequences of two functionally homologous enhancers are highly different, such that there is not enough similarity between them for sequence alignment programs to align them. By default, one would expect that these enhancers with highly dissimilar sequences to have equally dissimilar activity as well. Surprisingly, an increasing number of case studies are showing that this is not necessarily true. Some functionally homologous enhancers have highly divergent sequences in different species, even though they drive essentially identical expression patterns (Romano and Wray 2003b; Fisher et al. 2006; Wratten et al. 2006; Hare et al. 2008c). For instance, there is minimal sequence similarity between the sequences of *even-skipped* enhancers from *D. melanogaster* and *Sepsis cynipsae,* but the expression patterns they drive in a common *trans* acting environment (*D. melanogaster*) is virtually identical (Hare et al. 2008c). As an early developmental gene, improper expression of *even-skipped* can be detrimental for the organism (Ludwig et al. 2005; Ludwig et al. 2011). Hence it is intriguing to find that throughout evolution a high number of mutations in an enhancer that is essential for the organism were passed onto the next generation without being selected against.

One interpretation of the above observation is that stabilizing selection conserves function despite sequence turnover. This can happen through the combination of two mechanisms.

Firstly, degeneracy of transcription factor binding sites can allow sequence turnover despite functional conservation in enhancers. As mentioned previously, this is to say, a transcription factor can bind to multiple sequence motifs that share only part of their sequence (Bulyk 2003). As a result not all sequence changes would change the identity of a transcription factor binding site.

Degenerate nature of transcription factor binding sites were first shown *in vitro*, however it is also *in vivo*. Odom and colleagues (Odom et al. 2007) compared binding profiles of four highly conserved transcription factors (FOXA2, HNF1A, HNF4A, and HNF6) in mouse versus human hepatocytes using chromatin immuno precipitation (CHIP) followed by microarrays (chip) for over 4,000 orthologous regulatory regions. The authors found that the location of the majority of the transcription factor binding events, with respect to the coding sequences, were divergent between the two species. Moreover, among the binding events detected close to the promoter of orthologous genes, sequences of approximately two thirds of the regions that were found to bind the same transcription factor failed to align between human and mouse. This result is consistent with a previous analysis, where a similar comparison was made using a smaller set of genes (51 genes) (Dermitzakis and Clark 2002).

Subsequent studies showed that the above results were not caused by changes in (the sequence preference of) the transcription factors. Wilson and colleagues (Wilson et al. 2008) compared the binding of three of the above transcription factors on human chromosome 21 placed in human versus mouse hepatocytes. They saw that despite the differences in the arrangement of the binding events of the three transcription factors on human and mouse chromosomes in their native *trans* environment, when the human chromosome was put in mouse environment the human binding events were recapitulated. This suggests that *trans* regulatory factors were conserved and that despite diverged binding site sequence, the mouse transcription factors were able to recognize and bind to the human binding sites. This is also consistent with previous observations that transcription factors evolve slowly (Carroll 2008) and the amino acid sequences of DNA binding domains tend to be conserved (Luscombe and Thornton 2002). This is important since changes in transcription factor activity can affect multiple genes and, hence, have a higher chance to cause detrimental effects on the phenotype than changes in *cis*-regulatory sequences (Brickman et al. 2001).

Secondly, sequence divergence between functionally homologous enhancers, despite functional conservation, can also occur through gradual compensatory gain and loss of transcription factor binding sites. For instance, it has been shown that loss of one type of binding site can be compensated by the gain of the same or a different type of binding site somewhere else in the enhancer without changing its activity (Swanson et al. 2011). This can happen through gradual nucleotide substitutions as well as small insertion/deletions. For instance Shirangi and colleagues (Shirangi et al. 2009) showed that the expression of *desatF* gene was conserved between *D. melanogaster*, *D. sechelia*, and *D. erecta,* but there were differences in the sequences of the corresponding functionally homologous enhancers*. D. melanogaster desatF* enhancer harbored eight hexamer repeats that were essential for its activity in *D. melanogaster*, but were lacking in the *desatF* enhancers of the other two species. Intriguingly, these hexamers were created through a series of small deletions. These deletions likely removed regions that were essential for the expression of *desatF* in the ancestral species, which was compensated by the formation of the hexamer motifs in *D. melanogaster*.

Similar to the enhancers of some putatively orthologous genes with conserved expression, enhancer elements of co-regulated genes within an organism can show a substantial amount of variation in the grammar of the transcription factor binding sites they harbor (Brown et al. 2007; Weirauch and Hughes 2010), even though they drive highly similar expression patterns. The first case, however, is thought to result from divergence of the sequence of a common ancestral enhancer, whereas in the latter case two enhancers start different, potentially both at the sequence and functional level, but change over time to attain the same activity through different mutational events. The second case is still important in showing that different combinations or arrangements of transcription factor binding sites can result in the same enhancer activity.

For instance, comparison of the transcription factor binding site motifs between the promoters of twenty four ribosomal genes, from *S. cerevisiae* or *Candida albicans,* showed that despite tight co-expression, the arrangement of the identified binding motifs differed substantially (Weirauch and Hughes 2010). Similarly, this was true for nineteen enhancers that drive co-expression of muscle genes in two different species of *Ciona, C. savignyi* and *C. intestinalis*. Brown and colleagues (Brown et al. 2007) showed that, despite strict co-expression in the embryonic tail muscle, the organization of the same three transcription factor binding sites is different in all nineteen enhancers of either species. In these two cases it is not known if there are additional sites other than the previously identified ones, however it is possible for compensatory mutations to give rise to binding sites for other appropriate transcription factors (i.e., that are expressed under the appropriate conditions for the gene they affect).

Given the importance of enhancers for organismal physiology and evolution, it is important to understand how enhancer sequence and activity changes over time. All of the above cases and interpretations suggest that enhancer activity can be conserved with or without being accompanied by sequence conservation. This is achieved thanks to the degeneracy of transcription factor binding sites and compensatory mutations that lead to gradual gain and loss binding sites. On the other hand, even few sequence changes can alter enhancer activity, through changes in transcription factor binding sites, as well as in their spacing and orientation or nucleosome positioning. Currently, functional assays are still the golden criteria for studying the effects of changes in enhancer activity.

How are enhancer activities gained and lost?

Just as mutations can lead to changes in existing enhancer activity they can also lead to the complete loss or gain of an enhancer element. Enhancer gain and loss is an important mechanism for evolution of gene expression because it typically leads to more drastic changes in expression than alterations of enhancer activities. For instance, loss of enhancer activity can lead to lack of gene expression in a tissue and developmental time where the gene used to be expressed. Similarly, a gene can gain an enhancer that drives expression in a domain where it did not used to be expressed before. These types of changes, arguably, may be more likely to affect phenotypes than altered activity of an existing enhancer.

So what are some of the different types of mutational mechanisms that can lead to gain or loss of enhancer activity? Below, I describe some of these mechanisms with examples.

How can an existing enhancer activity be lost?

The first mechanism that comes to mind for loss of enhancer activity is mutations that inactivate one or more essential transcription factor binding sites in an enhancer resulting in loss of its function. This was exemplified by Pan and colleagues (Pan et al. 2001), where the authors showed that nucleotide substitutions in an essential transcription factor binding site, cyclic AMP responsive element (CRE), in the mouse *Ren-1c* enhancer disrupts binding of the corresponding transcription factor (CREB/CREM), resulting in complete loss of enhancer activity under cell culture conditions. Similarly, Sporn and colleagues (Sporn and Schwarzbauer 1995) inactivated essential transcription factor binding sites in the fibroblast enhancer of the *fibronectin* (FN) gene by deleting a 27 bp-region that harbors them. This eliminated the expression of the *FN* gene fibroblast cells. Also, *in vivo* (transgenic *D. melanogaster)*, Jeong and colleagues (Jeong et al. 2006a) showed that only a total of seven nucleotide substitutions in two essential binding sites, both for Abdominal-B, results in loss of Abd-B binding as well as subsequent loss of male specific abdominal activity of the *yellow* body enhancer.

Another mechanism, perhaps not as common as inactivation of transcription factor binding sites, is a deletion that is big enough to remove a whole enhancer and hence its activity. Such an event can presumably take place during cellular processes such as chromosomal rearrangements, DNA repair through non homologous end joining, or imprecise transposon movement. Regardless of the molecular mechanism, case studies show that both evolution and diseases have taken advantage of deletion of enhancers.

For instance, Chan et al. (Chan et al. 2010) showed that a 501 bp 5' sequence, which is sufficient to drive *Pituitary homeobox transcription factor 1* (*Pitx1*) expression in the pelvic region of transgenic three spine sticklebacks, when deleted abolished enhancer activity as well as formation of pelvic spines in the corresponding body part. This deletion in fact makes up the majority of the genetic basis lying behind lack of pelvic spines in freshwater sticklebacks as compared to the marine populations.

In terms of deletions of enhancers resulting in disease states, Loots and colleagues (Loots et al. 2005) discovered that a 52 kb 5' element, missing in Van Buchem patients and is genetically linked to this disease, is responsible for driving sclerostin (*SOST* -- a negative regulator of bone formation) expression in the rib, skull, and femur, but not in kidney or heart in transgenic mice. This showed that this region harbored a bone specific enhancer, lack of which is potentially responsible for Van Buchem disease, a disorder that leads to progressive increase in bone density (Wergedal et al. 2003).

There are also studies showing deletion of tissue specific enhancers at a genomic level. For instance, McLean et al. (McLean et al. 2011) found over 500 genomic regions that are conserved between chimp and macaque, but are missing in humans. These conserved (between chimp and macaque) and deleted (in humans) regions (CONDELs) have a median size of 2,804 bp and the authors showed that all except for one of these CONDELs correspond to non-coding DNA elements, suggesting that these deleted regions may harbor *cis*-regulatory activity . This was further supported by the fact that two of the CONDELs showed tissue specific enhancer activity in transgenic mice, in a pattern that correlated with the expression pattern of the genes (*androgen receptor* - *AR*) nearby. Expression of these genes are thought to be important for formation of sensory vibrissae and penile spines in mice, which are anatomical features lost in the human lineage.

Other possible alternative mechanisms that can lead to loss of enhancer activity include insertions of various sizes that can disrupt an existing transcription factor binding site rendering it nonfunctional and insertions and deletions of big or small size that can inactivate an enhancer not by disrupting or removing transcription factor binding sites, but by changing the spacing between them. This changes the grammar of the enhancer module, which can, for instance, affect the protein-protein interactions between transcription factors necessary for proper functioning of enhancers.

How can an enhancer activity be gained?

There are several different ways a gene can gain a new enhancer activity: *de novo* evolution, change in enhancer-promoter interactions, chromosomal rearrangements, transposable element insertion and co-option. The mutational mechanisms underlying these different process, however, can be the same. Below I describe each of the above listed ways of enhancer gain, with examples whenever possible, and list some of the possible mutational mechanisms that can facilitate those processes.

A certain genomic region lacking enhancer activity can gain this function *de novo* through an accumulation of sequence substitutions, insertions and deletions creating a functional sequence environment for an enhancer, e.g., a combination of transcription factor binding sites that will be bound by tissue specific transcription factors, which will recruit appropriate cofactors and act synergistically to alter the expression of a gene in a time, place and quantity specific manner. Even though, in theory, over time mutations can create such a functional regulatory environment (Stone and Wray 2001), to date, there is only one empirical examples of *de novo* gain of enhancer activity that I am aware of. Eichenlaub and Ettwiller (Eichenlaub and Ettwiller 2011) made use of the vestigial coding sequences that in teleosts, as compared to mammals, lost their coding ability following whole genome duplication, and through few sequence changes evolved into enhancers that drive part of the total expression pattern of their flanking genes. The ancestors of these *de novo* enhancers (that preserved coding ability) did not appear to have enhancer activity in teleost or mammals confirming the novelty of these *cis*-regulatory elements. This type of enhancer evolution may be hard to detect in other cases, because, for instance, in order to state that a certain enhancer was gained *de novo* in one species, this or any other enhancer activities should not be found in the orthologous region in its ancestor. Since the ancestor of a species cannot be investigated, one can look at the current relatives of the species to see if the orthologous genomic regions in those species have any kind of enhancer activity. This involves doing a comprehensive functional analyses of this orthologous region from multiple species, at various developmental stages (Rebeiz et al. 2011a), which is quite laborious and the stopping point is not clear. Hence, there are not even many case studies that conducted such an analyses and proved against *de novo* gain of enhancer activity.

Another mechanism for gain of enhancer activity can occur through changes in promoter-enhancer interactions. An enhancer and a promoter that did not used to communicate due to incompatibilities (e.g., existence/absence of TATA box or initiator sequences in the promoter) or existence of an insulator element between them, can start interacting if appropriate mutations occur in and around the enhancer and/or the promoter sequences such that they allow communication between the enhancer and the promoter. These mutations can be nucleotide substitutions, insertions or deletions that, for instance, convert a promoter without a TATA box to one with TATA box, which would allow this promoter and its corresponding gene be able to interact with enhancers that they did not used to communicate with. Else, mutations can inactivate an insulator between an enhancer and a promoter that did not used to interact with each other, allowing the enhancer to communicate with the promoter.

Changes in promoter-enhancer interaction can also occur due to chromosomal rearrangements, such as inversions, which can relocate *cis*-regulatory elements, such that an enhancer specific to one gene can have access to another gene. Cande and colleagues (Cande et al. 2009c) documented such an inversion. They showed that the early promoter of the *D. melanogaster ladybird* gene also has insulator activity such that the 3' cardiac enhancer of the gene, which drives dorsal mesoderm expression in the embryos, cannot affect the expression of *ladybird*'s 5' neighbor, *C15* gene. The authors found that, the *ladybird* locus is inverted in a distantly related species, *Tribolium castaneum* (the flour beetle), such that the insulator element is not located between the cardiac enhancer and the *C15* gene anymore. This allows the cardiac enhancer to control the expression of *C15*, and the *C15* gene to gain a new enhancer.

Even in the absence of enhancer-promoter incompatibilities, chromosomal rearrangements can relocate an enhancer or a gene such that the gene comes under the control of a new enhancer. For instance, Lettice and colleagues (Lettice et al. 2011) showed that in a patient with holoprosencephaly spectrum syndrome, *sonic hedgehog* (*Shh*) gene came under the control of a new enhancer due to an intrachromosomal inversion. This resulted in the enhancer driving ectopic expression of the gene in the developing limb bud, which leads to footplate expansion and eventual polydactyly in transgenic mice.

Transposon insertions provide another way through which genes can gain new enhancers. This typically requires the transposable element (TE) to have autonomous enhancer activity, but it can also function in combination with the flanking sequence of its insertion site. In the case (Daborn et al. 2002) of *D. melanogaster* cytochromoe P450 gene *Cyp6g1*, insertion of the long terminal repeat of the *Accord* retrotransposon 5' to the gene's coding sequence, resulted in highly upregulated expression of the gene in larval tissues that show basal level expression in the lack of the retrotransposon. These tissues are the midgut, Malpigian tubules and the fat body, all of which are important for detoxification in

Drosophila. Increased expression of *Cyp6g1* in these tissues also brings about insecticide resistance to the fly.

A similar example is from plants. Data collected by Studer and colleagues (Studer et al. 2011) strongly suggests that the insertion of the *Hopscotch* retrotransposon upstream of the *teosinte branched 1* (*tb1*) gene, which encodes a transcription factor involved in growth repression, is responsible for the two-fold increase in the gene’s expression in the axillary branches in maize versus its wild ancestor teosinte. Upregulated expression of *tb1* in this tissue is thought to be responsible for the reduced growth of branches in maize as compared to teosinte.

It is important to note that in the above two examples, the “gain of enhancer” was not expression of the gene at a new domain or developmental time, but rather a drastic increase in quantity, such that this led to a change in organismal phenotype. Overall, the mechanism of gaining an enhancer via transposon insertion is limited by the enhancer activities transposons can have as well as by their preferences on where to insert in the genome.

Co-option is another mechanism through which genes can gain new enhancers and be expressed in new tissue types. It requires a sequence fragment with an already existing enhancer activity to gain new transcription factor binding sites which work in collaboration with the already existing binding sites in the region and drive expression of the corresponding gene in a new domain. This new domain is determined by where the new binding factors are expressed. Co-option of enhancer function typically takes only a few mutations (Rebeiz et al. 2011a) and hence is suggested to be the most common mechanism for gain of enhancer activity. Consistent with this expectation, there is a growing number of case studies showing that gain of few transcription factor binding sites in and around an existing enhancer can lead to enhancer activity in new domains.

One of the best shown examples of co-option of enhancer function comes from the Drosophila *yellow* gene. A 5' wing enhancer drives the expression of this gene at low level throughout the pupal wing in both male and female *D. melanogaster*. However, in *D. biarmipes,* in addition to the dim expression throughout the pupal wing, the same region also drives elevated expression in the anterior distal wing spot, where *D. biarmipes* flies show elevated pigmentation as well. This high expression in the wing spot is due to a handful of nucleotide substitutions in the *yellow* wing enhancer of *D. biarmipes* as compared to the orthologous enhancer in *D. melanogaster* (and *D. pseudoobscura*). Some of these causative nucleotide changes led to gain of binding sites for the transcription factor Engrailed in *D. biarmipes* wing-spot enhancer, which has an expression domain that encompasses the anterior distal region of the wing (Gompel et al. 2005a), and is responsible for part of the observed wing-spot expression of *yellow*.

Another detailed study documenting co-option of enhancer activity investigated expression of the *Neprilysin1*(*Nep1*) gene among Drosophila species. Rebeiz and colleagues (Rebeiz et al. 2011a) found that this gene is expressed only in the mushroom bodies in the third instar larval brains in all nine species investigated except for *D. santomea*. In *D. santomea*, in addition to the mushroom bodies, *Nep1* is expressed also in the laminar neuroblasts of the optic lobe. They found that this new expression domain was acquired through only few (four) nucleotide substitutions that led to gain of new transcription factor binding sites neighboring the enhancer that drives expression in the central nervous system (CNS) and retinal field. The newly gained transcription factor binding sites worked together with the CNS-retinal field enhancer to drive expression in the optic lobe. In other words, the CNS-retinal field enhancer of *Nep1* was co-opted to gain novel activity in the optic lobe.

All in all, both gain and loss of enhancer activity are important evolutionary trajectories that can lead to phenotypic changes. It is intriguing that even though in general loss of enhancer activity is likely to lead to disruptive phenotypes, as exemplified by the resulting disease states, over evolutionary time some of the losses of enhancer activities brought about adaptive phenotypes. It is also intriguing that gain of enhancer activity, despite bearing the default expectation of requiring numerous mutational steps to get to, can actually be achieved even through few nucleotide substitutions. These examples overall show that enhancer structure and activity can be dynamic.

How does the *cis*-regulatory architecture of a gene change over time?

Genomic position of enhancers and how it changes over time has not been widely studied, and in fact, there is only few case studies focused on evolution of enhancer position specifically. Nevertheless, with respect to the coding sequences, enhancer position is generally thought to be conserved between species. That is to say, if an enhancer is located in the 5' intergenic region of a gene in one species it is expected to be found in the physically homologous genomic region, and not, for example, in the intron or the 3' intergenic region of the gene or in a different chromosome, in other species. This conservation of enhancer position has been documented either as a main (Cande et al. 2009c) or a side result (Hare et al. 2008a) in a number of case studies looking at orthologous enhancers.

For instance, Cande and colleagues (Cande et al. 2009c) found that the relative genomic positions of Dorsal target enhancers of several genes (*twist*, *brinker*, *cactus*, *single-minded*, *short-gastrulation*, *ventral nervous system defective*), which are necessary for dorsoventral patterning, were conserved through long evolutionary distances. The authors identified clusters of Dorsal binding sites around the target genes from the mosquito *Anopheles gambiae* and the flour beetle *Tribolium castaneum*. Reporter gene assays showed that regions harboring clusters of Dorsal binding sites from *Anopheles* and *Tribolium* harbor homologous enhancer activities to their *D. melanogaster* counterparts, i.e., they drive expression in the same developmental domain and time, despite few changes in specific expression patterns and lack of sequence similarity. Intriguingly, the authors also found that despite the long evolutionary distances of Anopheles and Tribolium from Drosophila (~250 mya (Yeates and Wiegmann 1999; Gaunt and Miles 2002) and 300 mya (Kristensen 1999), respectively) at least one of the enhancers for each gene identified in the former two species were located in the same position in the corresponding genomes, as they are in the *D. melanogaster* genome, with respect to the coding sequences of the cognate and neighboring genes in that region.

Consistent with the above study, Hare and colleagues (Hare et al. 2008a) documented, this time in only one locus, that the genomic organization of the four enhancers regulating expression of the *even-skipped* gene were conserved between several Drosophila and Sepsis species, split approximately 100 million years ago. This conservation of enhancer location was matched with conservation of activity, despite the high level of sequence divergence between enhancers from Drosophila and Sepsis species.

One can think of several reasons for why enhancer position, with respect to the coding sequences, would be conserved over time. Firstly, enhancers that regulate the expression of a cognate gene in the same manner and have the same genomic positions in different species can be orthologous, i.e., descendants of the same ancestral enhancer. In this case the reason for why their genomic positions, with respect to the coding sequences, are conserved may simply be because not enough number of appropriate mutations accumulated to change the location of the enhancer in one or more species. Here, an appropriate mutation would be one that contributes to altering the location of the particular enhancer, but does not negatively affect the fitness of the organism. Hence, in theory, any of the mutational mechanisms that can cause enhancer gain and loss can also lead to change in enhancer position as long as they are not disruptive to the organism.

Without excluding the above reasoning, one can think of a second explanation for conserved enhancer position. This explanation can involve orthologous or independently evolved enhancers that regulate a particular expression pattern of the same gene in different species. Basically, it is possible that at least for some genes, genomic position of the enhancers, with respect to the coding sequences, is essential in coordinating proper expression and changes in the position of an enhancer can potentially disrupt enhancer activity. In this case, mutations that alter the position of an enhancer in the genome would be selected against over time. This is in contrast to the textbook definition of enhancers, which states that enhancers show the same activity on their cognate gene independent of their position and orientation in the genome relative to the coding sequences. Even though some of the earlier case studies show that this assumption may be true for some enhancers (Banerji et al. 1981), it has not been directly tested until recently.

Intriguingly, Swanson et al. (Swanson et al. 2010) showed, in fine detail, that certain enhancer elements are position dependent for proper activity. The *sparkling* (*spa*) enhancer, which drives cone cell specific expression of the *dpax2* gene, has a “remote control element” that functions properly only when located at a distance from the basal promoter. Lack of this remote control element leads to loss of enhancer activity in cone cells when *spa* is placed far from the basal promoter (800bp). However, when *spa* is located in close proximity (121bp) to the basal promoter, lack of this element does not affect expression in cone cells. This shows that the proper functioning of the *spa* enhancer at a distance from the transcription start site is dependent on the remote control element.

Functional restrictions as above can lead evolutionary forces to conserve the genomic location of an enhancer element. However, currently we do not know how often enhancers have a position dependent components to them and how often there will be relaxed constrain on their positions.

Thirdly, observed conservation of enhancer position between species can be a by product of ascertainment bias. When scientists identify an enhancer and are looking for it in different species, they typically look at the physically homologous genomic sequence for high similarity to the already identified enhancer sequence. If they cannot find sequence similarity in the corresponding genomic region, they typically either interpret this as lack of the specific enhancer in that species or simply not report existence or absence of the enhancer. This leads to a level of ascertainment bias towards the reported cases where *cis*-regulatory architecture is conserved. More thorough research studies conduct functional assays to look whether the physically homologous regions of an enhancer in other species also have the same enhancer activity. This helps identify enhancers, the relative genomic locations of which are conserved, but are lacking sequence similarity. However, if a genomic region, physically homologous to a verified enhancer in another species, lacks enhancer activity then the possibility of change in enhancer location is rarely considered. These types of negative results also tend to not be included in publications, hence contributing to the possible ascertainment bias against altered *cis*-regulatory architectures.

Occasionally, however, researchers look for sequence conservation (Sanges et al. 2006) or a certain binding site grammar (Pan et al. 1994b) at other possible genomic regions in and around the corresponding gene. This type of search helpsed researchers find *cis*-regulatory elements that are organized differently in different species. For instance, Pan et colleagues (1994), showed that the proximal enhancer of the twist gene that drive expression in embryonic mesoderm, is located in the 5’ intregenic region of the coding sequence in *D. melanogaster*, but not in *D. virilis*. Search for the specific binding site grammar of this enhancer, identified in *D. melanogaster*, in and around the *twist* gene from *D. virilis*, found an enhancer in the intron that is functionally equivalent to the proximal enhancer. This indicated that the proximal enhancer of *twist* has different locations in different species.

Even in the absence of identifying similar sequences or binding site grammars, one needs to test candidate regions in and around a gene for enhancer activity to reveal a complete picture of the overall *cis*-regulatory architecture of a gene.

Dynamic structure and activity of *yellow cis*-regulatory sequences can be used as a model to understand how enhancers change over time

yellow expression patterns are highly divergent among Drosophila species, and many of these divergent patterns are attributed to differences in the *cis*-regulatory architecture (sequence, transcription factor binding profile, activity, position) of the gene (Wittkopp et al. 2002c; Gompel et al. 2005a; Jeong et al. 2006a; Prud'homme et al. 2006a). This diversity provides a great opportunity for addressing a long standing questions in the field: How do enhancers evolve? More specifically, how do their sequence, transcription factor binding profile, genomic organization, and activity change over time and how do these three features correlate with each other?

What is the function of the yellow *gene?*

yellow was first identified in *D. melanogaster* as a gene required for black pigment (melanin) formation in the larval mouthparts and the adult body and bristles (Morgan and Bridges 1916; Brehme 1941; Biessmann 1985). Later on, it was also found to be necessary for proper male courtship behavior (specifically for the wing extension part of the male courtship ritual) (Bastock 1956; Burnet et al. 1973; Drapeau et al. 2003).

The molecular function of the Yellow protein is still not known, however there are studies suggesting various different activities for it. For instance, temporal and spatial expression of *yellow* gene is in correlation with cuticle formation and dopa decarboxylase enzyme activity in Drosophila development (Walter et al. 1991).

However, in an insect cell/baculovirus expression system Yellow failed to show Dopachrome-conversion enzyme (DCE) activity, which accelerates insect melanization reactions significantly (Han et al. 2002). Based on its sequence, it is also thought to be a secreted protein (Geyer et al. 1986).

Based on sequence similarity, there are 13 additional *yellow*-like genes found in *D. melanogaster* (*yellow-b, -c, -d, - d2, -e, -e2, -e3, -f, -f2, -g, -h, -k*) (Maleszka and Kucharski 2000; Drapeau 2001). Among these Yellow-f and Yellow-f2 were found to have DCE activity in an insect cell/ baculovirus expression system (Han et al. 2002). *yellow* and *yellow*-like genes have been found to contain a shared domain with “major royal jelly” proteins (major royal jelly protein domain - MRJP), which are required for eusocial behavior in honey bees (*Apis mellifera*). As a result of searches based on the existence of MRJP domain there were no *yellow*-like genes found in non-insect metazoans, except for the putative sequences in a basal eukaryote (Naegleria), a primitive chordate (Amphioxus), and a crustecean (the Salmon Louse) (Ferguson et al. 2011). Interestingly *yellow*-like genes have been identified in 45 different species of bacteria as well as both ascomycete and basidiomycete fungal species (Maleszka and Kucharski 2000; Drapeau et al. 2006*a*; Ferguson et al. 2011), which led to speculations about horizontal gene transfer. *yellow*-like genes are also common in different species of insects. So far, *yellow*-like genes have been identified from *Apis mellifera* (20 genes, (Drapeau et al. 2006a)), *Bombyx mori* (seven genes, (Xia et al. 2006)), *Tribolium castaneum* (14 genes,), *Nasonia vitripennis* (25 genes, (Werren et al. 2010)), pea aphid *Acyrthosiphon pisum* (14 genes, (Ferguson et al. 2011)). A phylogenetic tree based on the sequence similarity of all inferred *yellow*-like proteins showed that *yellow* family expansion took place through gene duplications and is associated with insect diversification (Ferguson et al. 2011).

In many Drosophila species (Wittkopp and Beldade 2009) as well as other insect species (*B. mori,* (Ito et al. 2010), *T. casteneum*,), *yellow* has been found to be associated to pigmentation of different body parts, including body cuticle and wings. In *Aedes aegypti* the protein product of a *yellow*-like gene was even found to have DCE activity (Johnson et al. 2001). Further studies will show the specific functions of the *yellow* gene as well as different members of its gene family.

How is yellow *regulated?*

In *D. melanogaster, yellow* has two exons and a lone intron. Previous studies showed that both 5' intergenic and intronic regions of this gene harbor enhancer activities; the former drives expression in the body (epidermal cells in the abdomen, thorax and head), and wings (wing epidermal and vein cells) (Geyer and Corces 1987b; Martin et al. 1989a; Wittkopp et al. 2002c; Gompel et al. 2005a) as well as cells in the 3rd instar larval CNS harbors (referred to as mating success regulatory sequence MRS - necessary for proper male mating success, (Drapeau et al. 2005)) whereas the latter in the bristle associated cells (Geyer and Corces 1987b; Martin et al. 1989a).

Expression patterns of *yellow* are highly divergent between species, mostly due to changes in its *cis*-regulatory elements. Changes in *yellow* expression in the body between *D. melanogaster*, *D. subobscura* and *D. virilis* (Wittkopp et al. 2002c), in the wings between *D. melanogaster*, *D. biarmipes* and *D. pseudoobscura* (Gompel et al. 2005a), in the wings between *D. elegans*, *D. gununcola*, *D mimetica* and *D. tristis* (Prud'homme et al. 2006a) have been (partially or fully) attributed to altered activity of *yellow* enhancers. It is notable, however, that the differences in the *yellow* wing expression patterns between *D. melanogaster* and *D. guttifera* were caused by changes in the *trans* environment of the two species (Werner et al. 2010a).

So far there are only a handful of transcription factors that have been shown to directly or indirectly affect *yellow* expression. Transcription factor Fruitless (Fru) (Drapeau et al. 2003), was found to be genetically upstream of *yellow* in D. melanogaster, such that it has upregulatory effects on Yellow expression. Other transcription factors, Abdominal-B (Abd-B) (Jeong et al. 2006a) and Engrailed (En) (Gompel et al. 2005a) have been shown to affect *yellow* expression directly by binding to its 5' *cis*-regulatory elements. Abd-B binding events were shown *in vivo* for *D. melanogaster yellow* 5' intergenic region and *in vitro* for *D. santomea*, *D. biarmipes*, *D. kikkawai*, *D. bipectinata*, *D. subobscura* (spanning ~25 my of evolutionary history) (Jeong et al. 2006a). En was shown to bind directly to 5' intergenic region of *yellow* in vivo in *D. biramipes* only (Gompel et al. 2005a). Despite the advantage the diversity of *yellow cis*-regulatory regions provides in understanding *cis*-regulatory evolution, limited knowledge on the transcription factors binding to those regions is a challenge to overcome, before addressing questions about how transcription factor binding profiles change over time and how those changes are correlated to changes in enhancer activities.

The dynamic structure and activity of *yellow* enhancers make evolution of *yellow cis*-regulatory architecture a good prospective model for understanding how enhancers change over time. A global comparison of *yellow cis*-regulatory architecture between species can shed light onto some of the general trends or trajectories evolution takes to alter enhancers (structure and/or activity). Intriguingly, even though there is considerable knowledge, even at the nucleotide resolution, about differences between the *cis*-regulatory regions of *yellow* among several groups of species, an extensive comparison of structure and activity of *yellow* enhancers among Drosophila species has not yet been done. Moreover, again, despite the current ample knowledge on *yellow cis*-regulation, there are only a handful of *trans* factors shown to directly or indirectly regulate *yellow* expression. This raises another long standing question in the gene regulation field: How do *cis*-regulatory changes correlate to *trans* changes, i.e., what is the relationship between the trio of *cis*-regulatory sequences, transcription factor binding profiles, and *cis*-regulatory activities?

In order to address the above-mentioned questions, and as described in Chapter 2, I first identified the body, wing and bristle associated cell enhancer activities lying in the 5’ intergenic and intronic regions of *yellow* from multiple species spanning the Drosophila evolutionary history. This also allowed me to determine the genomic organization of these enhancers with respect to the *yellow* coding region. Subsequently, as described in Chapter 3, in order to find how the identified enhancer activities were partitioned within the full *yellow* 5’ intergenic and intronic regions, i.e., what types of enhancer activities came together to make up the expression pattern the full region drives, I tested sub-fragments of the full regions for enhancer activity. With the goal of investigating the relationship between differences in enhancer activities and differences in the corresponding transcription factor binding profiles, I used a yeast-one-hybrid screen and found, at least part of, the transcription factors that bind to *yellow* enhancer sub-elements, some of which had strong similarities or differences in their enhancer activities. Based on this work, and as described in Chapter 4, I propose that under appropriate conditions enhancer position and activity can change rapidly and understanding the underlying mechanism of this change requires a comprehensive analysis of how similarities and differences in sequence and transcription factor binding grammars correlate to or cause similarities and differences between enhancer activities.

References

Andrews, H. K., N. Giagtzoglou, S. Yamamoto, K. L. Schulze, and H. J. Bellen. 2009. Sequoia regulates cell fate decisions in the external sensory organs of adult Drosophila. EMBO reports 10:636-641.

Arakane, Y., N. T. Dittmer, Y. Tomoyasu, K. J. Kramer, S. Muthukrishnan, R. W. Beeman, and M. R. Kanost. Identification, mRNA expression and functional analysis of several yellow family genes in Tribolium castaneum. Insect Biochem Mol Biol 40:259-266.

Baker, B. S., and K. A. Ridge. 1980. Sex and the single cell. I. On the action of major loci affecting sex determination in Drosophila melanogaster. Genetics 94:383-423.

Baker, B. S., B. J. Taylor, and J. C. Hall. 2001. Are complex behaviors specified by dedicated regulatory genes? Reasoning from Drosophila. Cell 105:13-24.

Banerji, J., S. Rusconi, and W. Schaffner. 1981. Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. Cell 27:299-308.

Barolo, S. 2012. Shadow enhancers: Frequently asked questions about distributed cis-regulatory information and enhancer redundancy. BioEssays 34:135-141.

Bastock, M. 1956. A Gene Mutation Which Changes a Behavior Pattern. Evolution 10:421-439.

Baumann, M., J. Pontiller, and W. Ernst. 2010. Structure and Basal Transcription Complex of RNA Polymerase II Core Promoters in the Mammalian Genome: An Overview. Molecular Biotechnology 45:241-247.

Bell, A. C., A. G. West, and G. Felsenfeld. 1999. The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. Cell 98:387-396.

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

Bischof, J., R. K. Maeda, M. Hediger, F. Karch, and K. Basler. 2007. An optimized transgenesis system for Drosophila using germ-line-specific phiC31 integrases. Proc Natl Acad Sci U S A 104:3312-3317.

Blackwood, E. M., and J. T. Kadonaga. 1998. Going the Distance: A Current View of Enhancer Action. Science 281:60-63.

Blanchette, M., W. J. Kent, C. Riemer, L. Elnitski, A. F. Smit, K. M. Roskin, R. Baertsch, K. Rosenbloom, H. Clawson, E. D. Green, D. Haussler, and W. Miller. 2004. Aligning multiple genomic sequences with the threaded blockset aligner. Genome Res 14:708-715.

Booth, L. N., B. B. Tuch, and A. D. Johnson. 2010. Intercalation of a new tier of transcription regulation into an ancient circuit. Nature 468:959-963.

Brehme, K. S. 1941. The effect of adult body color mutations upont he larva of Drosophila melanogaster. Proceedings of the National Academy of Sciences 27:254-261.

Brenman, J. E., F. B. Gao, L. Y. Jan, and Y. N. Jan. 2001. Sequoia, a tramtrack-related zinc finger protein, functions as a pan-neural regulator for dendrite and axon morphogenesis in Drosophila. Developmental Cell 1:667-677.

Brickman, J. M., M. Clements, R. Tyrell, D. McNay, K. Woods, J. Warner, A. Stewart, R. S. Beddington, and M. Dattani. 2001. Molecular effects of novel mutations in Hesx1/HESX1 associated with human pituitary disorders. Development 128:5189-5199.

Britten, R. J., and E. H. Davidson. 1971. Repetitive and non-repetitive DNA sequences and a speculation on the origins of evolutionary novelty. The Quarterly review of biology 46:111-138.

Brown, C. D., D. S. Johnson, and A. Sidow. 2007. Functional Architecture and Evolution of Transcriptional Elements That Drive Gene Coexpression. Science 317:1557-1560.

Bulger, M., and M. Groudine. 2011. Functional and mechanistic diversity of distal transcription enhancers. Cell 144:327-339.

Bulyk, M. L. 2003. Computational prediction of transcription-factor binding site locations. Genome Biology 5:201.

Burnet, B., K. Connolly, and B. Harrison. 1973. Phenocopies of pigmentary and behavioral effects of the yellow mutant in Drosophila induced by alpha-dimethyltyrosine. Science 181:1059-1060.

Butler, J. E., and J. T. Kadonaga. 2001. Enhancer-promoter specificity mediated by DPE or TATA core promoter motifs. Genes & Development 15:2515-2519.

Cande, J., Y. Goltsev, and M. Levine. 2009a. Conservation of enhancer location in divergent insects. Proc Natl Acad Sci U S A 106:14414-14419.

Cande, J., Y. Goltsev, and M. S. Levine. 2009b. Conservation of enhancer location in divergent insects. Proc Natl Acad Sci U S A 106:14414-14419.

Cande, J. D., V. S. Chopra, and M. Levine. 2009c. Evolving enhancer-promoter interactions within the tinman complex of the flour beetle, Tribolium castaneum. Development 136:3153-3160.

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

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

Chan, Y. F., M. E. Marks, F. C. Jones, G. Villarreal, M. D. Shapiro, S. D. Brady, A. M. Southwick, D. M. Absher, J. Grimwood, J. Schmutz, R. M. Myers, D. Petrov, B. Jónsson, D. Schluter, M. A. Bell, and D. M. Kingsley. 2010. Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer. Science 327:302-305.

Cléard, F., and P. Spierer. 2001. Position-effect variegation in Drosophila: the modifier Su(var)3-7 is a modular DNA-binding protein. EMBO reports 2:1095-1100.

Couderc, J. L., D. Godt, S. Zollman, J. Chen, M. Li, S. Tiong, S. E. Cramton, I. Sahut-Barnola, and F. A. Laski. 2002. The bric a brac locus consists of two paralogous genes encoding BTB/POZ domain proteins and acts as a homeotic and morphogenetic regulator of imaginal development in Drosophila. Development 129:2419-2433.

Cretekos, C. J., Y. Wang, E. D. Green, J. F. Martin, J. J. t. Rasweiler, and R. R. Behringer. 2008. Regulatory divergence modifies limb length between mammals. Genes Dev 22:141-151.

Daborn, P. J., J. L. Yen, M. R. Bogwitz, G. Le Goff, E. Feil, S. Jeffers, N. Tijet, T. Perry, D. Heckel, P. Batterham, R. Feyereisen, T. G. Wilson, and R. H. ffrench-Constant. 2002. A single p450 allele associated with insecticide resistance in Drosophila. Science 297:2253-2256.

de Celis, J. F., M. Llimargas, and J. Casanova. 1995. Ventral veinless, the gene encoding the Cf1a transcription factor, links positional information and cell differentiation during embryonic and imaginal development in Drosophila melanogaster. Development 121:3405-3416.

de Hoon, M. J. L., S. Imoto, J. Nolan, and S. Miyano. 2004. Open source clustering software. Bioinformatics 20:1453-1454.

Déjardin, J., and R. E. Kingston. 2009. Purification of proteins associated with specific genomic Loci. Cell 136:175-186.

Dekker, J., K. Rippe, M. Dekker, and N. Kleckner. 2002. Capturing chromosome conformation. Science 295:1306-1311.

Dermitzakis, E., and A. Clark. 2002. Evolution of Transcription Factor Binding Sites in Mammalian Gene Regulatory Regions: Conservation and Turnover. Mol Biol Evol 19:1114-1121.

Diaz-Benjumea, J., M. A. F. Gonzalez Gaitan, and A. Garcia-Bellido. 1989. Developmental genetics of the wing vein pattern of *Drosophila*. Genome 31:612-619.

Drapeau, M. 2001. The Family of Yellow-Related Drosophila melanogaster Proteins. Biochemical and Biophysical Research Communications 281:611-613.

Drapeau, M. 2006. FRU does not regulate yellow in fly pigmented cells. PHILICA.COM Observation number 13.

Drapeau, M. D., S. Albert, R. Kucharski, C. Prusko, and R. Maleszka. 2006a. Evolution of the Yellow/Major Royal Jelly Protein family and the emergence of social behavior in honey bees. Genome Research 16:1385-1394.

Drapeau, M. D., S. A. Cyran, M. M. Viering, P. K. Geyer, and A. D. Long. 2005. A cis-regulatory Sequence Within the yellow Locus of Drosophila melanogaster Required for Normal Male Mating Success. Genetics 172:1009-1030.

Drapeau, M. D., S. A. Cyran, M. M. Viering, P. K. Geyer, and A. D. Long. 2006b. A cis-regulatory sequence within the yellow locus of Drosophila melanogaster required for normal male mating success. Genetics 172:1009-1030.

Drapeau, M. D., A. Radovic, P. J. Wittkopp, and A. D. Long. 2003. A gene necessary for normal male courtship, yellow, acts downstream of *fruitless* in the *Drosophila melanogaster* larval brain. Journal of Neurobiology 55:53-72.

Dubruille, R., A. Laurençon, C. Vandaele, E. Shishido, M. Coulon-Bublex, P. Swoboda, P. Couble, M. Kernan, and B. Durand. 2002. Drosophila Regulatory factor X is necessary for ciliated sensory neuron differentiation. Development 129:5487-5498.

Durand, B., C. Vandaele, D. Spencer, S. Pantalacci, and P. Couble. 2000. Cloning and characterization of dRFX, the Drosophila member of the RFX family of transcription factors. Gene 246:285-293.

Eichenlaub, M. P., and L. Ettwiller. 2011. De Novo Genesis of Enhancers in Vertebrates. PLoS Biology 9:e1001188.

Elgar, G., and T. Vavouri. 2008. Tuning in to the signals: noncoding sequence conservation in vertebrate genomes. Trends in Genetics 24:344-352.

Ferguson, L. C., J. Green, A. Surridge, and C. D. Jiggins. 2011. Evolution of the Insect Yellow Gene Family. Molecular Biology and Evolution 28:257-272.

Fisher, S., E. A. Grice, R. M. Vinton, S. L. Bessling, and A. S. McCallion. 2006. Conservation of RET regulatory function from human to zebrafish without sequence similarity. Science 312:276-279.

Fisk, G. J., and C. S. Thummel. 1998. The DHR78 Nuclear Receptor Is Required for Ecdysteroid Signaling during the Onset of Drosophila Metamorphosis. Cell 93:543-555.

Frankel, N., G. K. Davis, D. Vargas, S. Wang, F. Payre, and D. L. Stern. 2010a. Phenotypic robustness conferred by apparently redundant transcriptional enhancers. Nature 466:490-493.

Frankel, N., G. K. Davis, D. Vargas, S. Wang, F. Payre, and D. L. Stern. 2010b. Phenotypic robustness conferred by apparently redundant transcriptional enhancers. Nature:May 30. [Epub ahead of print].

Frankel, N., D. F. Erezyilmaz, A. P. McGregor, S. Wang, F. Payre, and D. L. Stern. 2011. Morphological evolution caused by many subtle-effect substitutions in regulatory DNA. Nature 474:598-603.

Futahashi, R., and H. Fujiwara. 2007. Regulation of 20-hydroxyecdysone on the larval pigmentation and the expression of melanin synthesis enzymes and yellow gene of the swallowtail butterfly, Papilio xuthus. Insect Biochemistry and Molecular Biology 37:855-864.

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

Gaunt, M., and M. Miles. 2002. An Insect Molecular Clock Dates the Origin of the Insects and Accords with Palaeontological and Biogeographic Landmarks. Mol Biol Evol 19:748-761.

Geyer, P. K., and V. G. Corces. 1987a. Separate regulatory elements are responsible for the complex pattern of tissue-specific and developmental transcription of the yellow locus in Drosophila melanogaster. Genes Devel. 1:996-1004.

Geyer, P. K., and V. G. Corces. 1987b. Separate regulatory elements are responsible for the complex pattern of tissue-specific and developmental transcription of the yellow locus in Drosophila melanogaster. Genes & Development 1:996-1004.

Geyer, P. K., C. Spana, and V. G. Corces. 1986. On the molecular mechanism of gypsy-induced mutations at the yellow locus of Drosophila melanogaster. The EMBO journal 5:2657-2662.

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

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

Goode, D. K., H. A. Callaway, G. A. Cerda, K. E. Lewis, and G. Elgar. 2011. Minor change, major difference: divergent functions of highly conserved cis-regulatory elements subsequent to whole genome duplication events. Development 138:879-884.

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

Gu, W., and R. G. Roeder. 1997. Activation of p53 sequence-specific DNA binding by acetylation of the p53 C-terminal domain. Cell 90:595-606.

Guo, H. F., J. Tong, F. Hannan, L. Luo, and Y. Zhong. 2000. A neurofibromatosis-1-regulated pathway is required for learning in Drosophila. Nature 403:895-898.

Hadzhiev, Y., M. Lang, R. Ertzer, A. Meyer, U. Strähle, and F. Muller. 2007. Functional diversification of sonic hedgehog paralog enhancers identified by phylogenomic reconstruction. Genome Biology 8:R106.

Han, Q., J. Fang, H. Ding, J. K. Johnson, B. M. Christensen, and J. Li. 2002. Identification of Drosophila melanogaster yellow-f and yellow-f2 proteins as dopachrome-conversion enzymes. The Biochemical Journal 368:333-340.

Hare, E. E., B. K. Peterson, and M. B. Eisen. 2008a. A Careful Look at Binding Site Reorganization in the even-skipped Enhancers of Drosophila and Sepsids. PLoS Genetics 4:e1000268.

Hare, E. E., B. K. Peterson, V. N. Iyer, R. Meier, and M. B. Eisen. 2008b. Sepsid even-skipped enhancers are functionally conserved in Drosophila despite lack of sequence conservation. PLoS Genet 4:e1000106.

Hare, E. E., B. K. Peterson, V. N. Iyer, R. Meier, and M. B. Eisen. 2008c. Sepsid even-skipped Enhancers Are Functionally Conserved in Drosophila Despite Lack of Sequence Conservation. PLoS Genetics 4:e1000106.

Heitzler, P., M. Haenlin, P. Ramain, M. Calleja, and P. Simpson. 1996. A Genetic Analysis of pannier, a Gene Necessary for Viability of Dorsal Tissues and Bristle Positioning in Drosophila. Genetics 143:1271-1286.

Helms, A. W., A. L. Abney, N. Ben-Arie, H. Y. Zoghbi, and J. E. Johnson. 2000. Autoregulation and multiple enhancers control Math1 expression in the developing nervous system. Development 127:1185-1196.

Hens, K., J. D. Feuz, A. Isakova, A. Iagovitina, A. Massouras, J. Bryois, P. Callaerts, S. E. Celniker, and B. Deplancke. 2011. Automated protein-DNA interaction screening of Drosophila regulatory elements. Nat Methods 8:1065-1070.

Hershberg, R., and D. A. Petrov. 2008. Selection on Codon Bias. Annual Review of Genetics 42:287-299.

Ho, I. S., F. Hannan, H. F. Guo, I. Hakker, and Y. Zhong. 2007. Distinct Functional Domains of Neurofibromatosis Type 1 Regulate Immediate versus Long-Term Memory Formation. Journal of Neuroscience 27:6852-6857.

Hong, J. W., D. A. Hendrix, and M. S. Levine. 2008. Shadow enhancers as a source of evolutionary novelty. Science 321:1314.

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

Horn, C., and E. A. Wimmer. 2000. A versatile vector set for animal transgenesis. Dev Genes Evol 210:630-637.

Ito, K., S. Katsuma, K. Yamamoto, K. Kadono-Okuda, K. Mita, and T. Shimada. 2010. Yellow-e Determines the Color Pattern of Larval Head and Tail Spots of the Silkworm Bombyx mori. Journal of Biological Chemistry 285:5624-5629.

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

Jeong, S., A. Rokas, and S. B. Carroll. 2006a. Regulation of Body Pigmentation by the Abdominal-B Hox Protein and Its Gain and Loss in Drosophila Evolution. Cell 125:1387-1399.

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

Jeong, Y., K. El-Jaick, E. Roessler, M. Muenke, and D. J. Epstein. 2006c. A functional screen for sonic hedgehog regulatory elements across a 1 Mb interval identifies long-range ventral forebrain enhancers. Development 133:761-772.

Johnson, J. K., J. Li, and B. M. Christensen. 2001. Cloning and characterization of a dopachrome conversion enzyme from the yellow fever mosquito, Aedes aegypti. Insect Biochemistry and Molecular Biology 31:1125-1135.

Kleinjan, D. A., and V. van Heyningen. 2005. Long-Range Control of Gene Expression: Emerging Mechanisms and Disruption in Disease. The American Journal of Human Genetics 76:8-32.

Kopp, A., R. K. Blackman, and I. Duncan. 1999. Wingless, Decapentaplegic and EGF Receptor signaling pathways interact to specify dorso-ventral pattern in the adult abdomen of Drosophila. Development 126:3495-3507.

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

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

Kozlova, T., G. Lam, and C. S. Thummel. 2009. DrosophilaDHR38 nuclear receptor is required for adult cuticle integrity at eclosion. Developmental Dynamics 238:701-707.

Kristensen, N. P. 1999. Phylogeny of endopterygote insects, the most successful lineage of living organisms. Eur J Entomology 96:237-253.

Langeland, J. A., and S. Carroll. 1993. Conservation of regulatory elements controlling hairy pair-rule stripe formation. Development 117:585-596.

Lee, M.-J., and D. A. Stephenson. 2007. Recent developments in neurofibromatosis type 1. Current opinion in neurology 20:135-141.

Lettice, L. A., S. Daniels, E. Sweeney, S. Venkataraman, P. S. Devenney, P. Gautier, H. Morrison, J. Fantes, R. E. Hill, and D. R. FitzPatrick. 2011. Enhancer-adoption as a mechanism of human developmental disease. Human Mutation 32:1492-1499.

Llimargas, M., and J. Casanova. 1997. ventral veinless, a POU domain transcription factor, regulates different transduction pathways required for tracheal branching in Drosophila. Development 124:3273-3281.

Lomvardas, S., G. Barnea, D. J. Pisapia, M. Mendelsohn, J. Kirkland, and R. Axel. 2006. Interchromosomal interactions and olfactory receptor choice. Cell 126:403-413.

Loots, G. G. 2008. Genomic identification of regulatory elements by evolutionary sequence comparison and functional analysis. Advances in Genetics 61:269-293.

Loots, G. G., M. Kneissel, H. Keller, M. Baptist, J. Chang, N. M. Collette, D. Ovcharenko, I. Plajzer-Frick, and E. M. Rubin. 2005. Genomic deletion of a long-range bone enhancer misregulates sclerostin in Van Buchem disease. Genome Research 15:928-935.

Ludwig, M. Z., Manu, R. Kittler, K. P. White, and M. Kreitman. 2011. Consequences of Eukaryotic Enhancer Architecture for Gene Expression Dynamics, Development, and Fitness. PLoS Genetics 7:e1002364.

Ludwig, M. Z., A. Palsson, E. Alekseeva, C. M. Bergman, J. Nathan, and M. Kreitman. 2005. Functional Evolution of a cis-Regulatory Module. PLoS Biology 3:e93.

Ludwig, M. Z., N. H. Patel, and M. Kreitman. 1998. Functional analysis of eve stripe 2 enhancer evolution in *Drosophila*: rules governing conservation and change. Development 125:949-958.

Lukowitz, W., C. Schroder, G. Glaser, M. Hulskamp, and D. Tautz. 1994. Regulatory and coding regions of the segmentation gene hunchback are functionally conserved between Drosophila virilis and Drosophila melanogaster. Mech Dev 45:105-115.

Luscombe, N. M., and J. M. Thornton. 2002. Protein-DNA interactions: amino acid conservation and the effects of mutations on binding specificity. Journal of molecular biology 320:991-1009.

Maerkl, S. J. 2011. Next generation microfluidic platforms for high-throughput protein biochemistry. Current Opinion in Biotechnology 22:59-65.

Maleszka, R., and R. Kucharski. 2000. Analysis of Drosophila yellow-B cDNA Reveals a New Family of Proteins Related to the Royal Jelly Proteins in the Honeybee and to an Orphan Protein in an Unusual Bacterium Deinococcus radiodurans. Biochemical and Biophysical Research Communications 270:773-776.

Malik, S., and R. G. Roeder. 2010. The metazoan Mediator co-activator complex as an integrative hub for transcriptional regulation. Nature Reviews Genetics 11:761-772.

Markstein, M., C. Pitsouli, C. Villalta, S. E. Celniker, and N. Perrimon. 2008. Exploiting position effects and the gypsy retrovirus insulator to engineer precisely expressed transgenes. Nat Genet 40:476-483.

Martin, M., Y. B. Meng, and W. Chia. 1989a. Regulatory elements involved in the tissue-specific expression of the yellow gene of Drosophila. Mol Gen Genet 218:118-126.

Martin, M., Y. B. Meng, and W. Chia. 1989b. Regulatory elements involved in the tissue-specific expression of the yellow gene of Drosophila. Mol. Gen. Genet. 218:118-126.

McGregor, A. P., V. Orgogozo, I. Delon, J. Zanet, D. G. Srinivasan, F. Payre, and D. L. Stern. 2007. Morphological evolution through multiple cis-regulatory mutations at a single gene. Nature 448:587-590.

McLean, C. Y., P. L. Reno, A. A. Pollen, A. I. Bassan, T. D. Capellini, C. Guenther, V. B. Indjeian, X. Lim, D. B. Menke, B. T. Schaar, A. M. Wenger, G. Bejerano, and D. M. Kingsley. 2011. Human-specific loss of regulatory DNA and the evolution of human-specific traits. Nature 471:216-219.

Meier, S., S. G. Sprecher, H. Reichert, and F. Hirth. 2006. ventral veins lacking is required for specification of the tritocerebrum in embryonic brain development of Drosophila. Mech Dev 123:76-83.

Morgan, T. H., and C. B. Bridges. 1916. Sex-linked inheritance in Drosophila. Publs Carnegie Instn 237:1-88.

Näär, A. M., B. D. Lemon, and R. Tjian. 2001. Transcriptional Coactivator Complexes. Annual Review of Biochemistry 70:475-501.

Nègre, N., C. D. Brown, P. K. Shah, P. Kheradpour, C. A. Morrison, J. G. Henikoff, X. Feng, K. Ahmad, S. Russell, R. A. H. White, L. Stein, S. Henikoff, M. Kellis, and K. P. White. 2010. A comprehensive map of insulator elements for the Drosophila genome. PLoS Genetics 6:e1000814.

Ni, J.-Q., M. Markstein, R. Binari, B. Pfeiffer, L.-P. Liu, C. Villalta, M. Booker, L. Perkins, and N. Perrimon. 2008. Vector and parameters for targeted transgenic RNA interference in Drosophila melanogaster. Nature Methods 5:49-51.

Ni, J. Q., L. P. Liu, R. Binari, R. Hardy, H. S. Shim, A. Cavallaro, M. Booker, B. D. Pfeiffer, M. Markstein, H. Wang, C. Villalta, T. R. Laverty, L. A. Perkins, and N. Perrimon. 2009. A Drosophila Resource of Transgenic RNAi Lines for Neurogenetics. Genetics 182:1089-1100.

O'Quin, K. E., D. Smith, Z. Naseer, J. Schulte, S. D. Engel, Y.-H. E. Loh, J. T. Streelman, J. L. Boore, and K. L. Carleton. 2011. Divergence in cis-regulatory sequences surrounding the opsin gene arrays of African cichlid fishes. BMC Evolutionary Biology 11:120.

Odom, D. T., R. D. Dowell, E. S. Jacobsen, W. Gordon, T. W. Danford, K. D. MacIsaac, P. A. Rolfe, C. M. Conboy, D. K. Gifford, and E. Fraenkel. 2007. Tissue-specific transcriptional regulation has diverged significantly between human and mouse. Nature Genetics 39:730-732.

Ohtsuki, S., M. Levine, and H. N. Cai. 1998. Different core promoters possess distinct regulatory activities in the Drosophila embryo. Genes & Development 12:547-556.

Ong, C.-T., and V. G. Corces. 2011. Enhancer function: new insights into the regulation of tissue-specific gene expression. Nature Reviews Genetics 12:284-293.

Orphanides, G., and D. Reinberg. 2002. A unified theory of gene expression. Cell 108:439-451.

Pan, D., S. A. Valentine, and A. J. Courey. 1994a. The bipartite D. melanogaster twist promoter is reorganized in D. virilis. Mech Dev 46:41-53.

Pan, D., S. A. Valentine, and A. J. Courey. 1994b. The bipartite D. melanogaster twist promoter is reorganized in D. virilis. Mechanisms of development 46:41-53.

Pan, L., T. A. Black, Q. Shi, C. A. Jones, N. Petrovic, J. Loudon, C. Kane, C. D. Sigmund, and K. W. Gross. 2001. Critical roles of a cyclic AMP responsive element and an E-box in regulation of mouse renin gene expression. The Journal of biological chemistry 276:45530-45538.

Pappu, K. S., E. J. Ostrin, B. W. Middlebrooks, B. T. Sili, R. Chen, M. R. Atkins, R. Gibbs, and G. Mardon. 2005. Dual regulation and redundant function of two eye-specific enhancers of the Drosophila retinal determination gene dachshund. Development 132:2895-2905.

Parnell, T. J., M. M. Viering, A. Skjesol, C. Helou, E. J. Kuhn, and P. K. Geyer. 2003. An endogenous suppressor of hairy-wing insulator separates regulatory domains in Drosophila. Proceedings of the National Academy of Sciences of the United States of America 100:13436-13441.

Pennacchio, L. A., N. Ahituv, A. M. Moses, S. Prabhakar, M. A. Nobrega, M. Shoukry, S. Minovitsky, I. Dubchak, A. Holt, K. D. Lewis, I. Plajzer-Frick, J. Akiyama, S. De Val, V. Afzal, B. L. Black, O. Couronne, M. B. Eisen, A. Visel, and E. M. Rubin. 2006. In vivo enhancer analysis of human conserved non-coding sequences. Nature 444:499-502.

Perry, M. W., A. N. Boettiger, J. P. Bothma, and M. Levine. 2010. Shadow enhancers foster robustness of Drosophila gastrulation. Current Biology 20:1562-1567.

Perry, M. W., A. N. Boettiger, and M. Levine. 2011. Multiple enhancers ensure precision of gap gene-expression patterns in the Drosophila embryo. Proc Natl Acad Sci U S A 108:13570-13575.

Peterson, B. K., E. E. Hare, V. N. Iyer, S. Storage, L. Conner, D. R. Papaj, R. Kurashima, E. Jang, and M. B. Eisen. 2009. Big Genomes Facilitate the Comparative Identification of Regulatory Elements. PLoS ONE 4:e4688.

Pospisilik, J. A., D. Schramek, H. Schnidar, S. J. F. Cronin, N. T. Nehme, X. Zhang, C. Knauf, P. D. Cani, K. Aumayr, J. Todoric, M. Bayer, A. Haschemi, V. Puviindran, K. Tar, M. Orthofer, G. G. Neely, G. Dietzl, A. Manoukian, M. Funovics, G. Prager, O. Wagner, D. Ferrandon, F. Aberger, C.-c. Hui, H. Esterbauer, and J. M. Penninger. 2010. Drosophila Genome-wide Obesity Screen Reveals Hedgehog as a Determinant of Brown versus White Adipose Cell Fate. Cell 140:148-160.

Prabhakar, S., F. Poulin, M. Shoukry, V. Afzal, E. M. Rubin, O. Couronne, and L. A. Pennacchio. 2006. Close sequence comparisons are sufficient to identify human cis-regulatory elements. Genome Research 16:855-863.

Prabhakar, S., A. Visel, J. A. Akiyama, M. Shoukry, K. D. Lewis, A. Holt, I. Plajzer-Frick, H. Morrison, D. R. FitzPatrick, V. Afzal, L. A. Pennacchio, E. M. Rubin, and J. P. Noonan. 2008. Human-specific gain of function in a developmental enhancer. Science 321:1346-1350.

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

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

Ragoczy, T., A. Telling, T. Sawado, M. Groudine, and S. T. Kosak. 2003. A genetic analysis of chromosome territory looping: diverse roles for distal regulatory elements. Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology 11:513-525.

Rebeiz, M., N. Jikomes, V. Kassner, and S. Carroll. 2011a. 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 108:10036-10043.

Rebeiz, M., N. Jikomes, V. Kassner, and S. Carroll. 2011b. Evolutionary origin of a novel gene expression pattern through co-option of the latent activities of existing regulatory sequences. Proc Natl Acad Sci U S A 108:10036-10043.

Reuter, G., M. Giarre, J. Farah, J. Gausz, A. Spierer, and P. Spierer. 1990. Dependence of position-effect variegation in Drosophila on dose of a gene encoding an unusual zinc-finger protein. Nature 344:219-223.

Rhead, B., D. Karolchik, R. M. Kuhn, A. S. Hinrichs, A. S. Zweig, P. A. Fujita, M. Diekhans, K. E. Smith, K. R. Rosenbloom, B. J. Raney, A. Pohl, M. Pheasant, L. R. Meyer, K. Learned, F. Hsu, J. Hillman-Jackson, R. A. Harte, B. Giardine, T. R. Dreszer, H. Clawson, G. P. Barber, D. Haussler, and W. J. Kent. The UCSC Genome Browser database: update 2010. Nucleic Acids Res 38:D613-619.

Rhead, B., D. Karolchik, R. M. Kuhn, A. S. Hinrichs, A. S. Zweig, P. A. Fujita, M. Diekhans, K. E. Smith, K. R. Rosenbloom, B. J. Raney, A. Pohl, M.

Pheasant, L. R. Meyer, K. Learned, F. Hsu, J. Hillman-Jackson, R. A. Harte, B. Giardine, T. R. Dreszer, H. Clawson, G. P. Barber, D. Haussler, and W. J. Kent. 2010. The UCSC Genome Browser database: update 2010. Nucleic Acids Res 38:D613-619.

Romano, L. A., and G. A. Wray. 2003a. Conservation of Endo16 expression in sea urchins despite evolutionary divergence in both cis and trans-acting components of transcriptional regulation. Development 130:4187-4199.

Romano, L. A., and G. A. Wray. 2003b. Conservation of Endo16 expression in sea urchins despite evolutionary divergence in both cis and trans-acting components of transcriptional regulation. Development 130:4187-4199.

Russell, S. R., G. Heimbeck, C. M. Goddard, A. T. Carpenter, and M. Ashburner. 1996. The Drosophila Eip78C gene is not vital but has a role in regulating chromosome puffs. Genetics 144:159-170.

Russo, C. A., N. Takezaki, and M. Nei. 1995. Molecular phylogeny and divergence times of drosophilid species. Mol Biol Evol 12:391-404.

Sambrook, J., and D. Russell. 2001. Molecular Cloning: a Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor.

Sanges, R., E. Kalmar, P. Claudiani, M. D'Amato, F. Muller, and E. Stupka. 2006. Shuffling of cis-regulatory elements is a pervasive feature of the vertebrate lineage. Genome Biology 7:R56.

Sauna, Z. E., and C. Kimchi-Sarfaty. 2011. Understanding the contribution of synonymous mutations to human disease. Nature Reviews Genetics 12:683-691.

Schneuwly, S., R. Klemenz, and W. J. Gehring. 1987. Redesigning the body plan of Drosophilaby ectopic expression of the homoeotic gene Antennapedia. Nature 325:816-818.

Schoenfelder, S., I. Clay, and P. Fraser. 2010. The transcriptional interactome: gene expression in 3D. Current Opinion in Genetics & Development 20:127-133.

Schwartz, S., W. J. Kent, A. Smit, Z. Zhang, R. Baertsch, R. C. Hardison, D. Haussler, and W. Miller. 2003. Human-mouse alignments with BLASTZ. Genome Res 13:103-107.

Senger, K., G. Armstrong, W. Rowell, J. Kwan, M. Markstein, and M. Levine. 2004. Immunity Regulatory DNAs Share Common Organizational Features in Drosophila. Molecular Cell 13:19-32.

Shin, J. T., J. R. Priest, I. Ovcharenko, A. Ronco, R. K. Moore, C. G. Burns, and C. A. MacRae. 2005. Human-zebrafish non-coding conserved elements act in vivo to regulate transcription. Nucleic Acids Research 33:5437-5445.

Shirangi, T. R., H. D. Dufour, T. M. Williams, and S. B. Carroll. 2009. Rapid evolution of sex pheromone-producing enzyme expression in Drosophila. PLoS Biology 7:e1000168.

Siepel, A., G. Bejerano, J. S. Pedersen, A. S. Hinrichs, M. Hou, K. Rosenbloom, H. Clawson, J. Spieth, L. W. Hillier, S. Richards, G. M. Weinstock, R. K. Wilson, R. A. Gibbs, W. J. Kent, W. Miller, and D. Haussler. 2005. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Res 15:1034-1050.

Smale, S. T., and J. T. Kadonaga. 2003. The RNA Polymerase II Core Promoter. Annual Review of Biochemistry 72:449-479.

Sotillos, S., and J. F. de Celis. 2006. Regulation of decapentaplegic expression during Drosophila wing veins pupal development. Mech Dev 123:241-251.

Spicer, G. S., and C. D. Bell. 2002. Molecular Phylogeny of the Drosophila virilis Species Group (Diptera: Drosophilidae) Inferred from Mitochondrial 12S and 16S Ribosomal RNA Genes. . Annuals of the Entomological Society of America 95:156-161.

Spierer, A., F. Begeot, P. Spierer, and M. Delattre. 2008. SU(VAR)3-7 Links Heterochromatin and Dosage Compensation in Drosophila. PLoS Genetics 4:e1000066.

Sporn, S., and J. Schwarzbauer. 1995. Identification of an enhancer involved in tissue-specific regulation of the rat fibronectin gene. Nucleic Acids Research 23:3335-3342.

Stone, J. R., and G. A. Wray. 2001. Rapid evolution of cis-regulatory sequences via local point mutations. Molecular Biology and Evolution 18:1764-1770.

Stormo, G. D. 2000. DNA binding sites: representation and discovery. Bioinformatics 16:16-23.

Stothard, P. 2000. The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. Biotechniques 28:1102, 1104.

Studer, A., Q. Zhao, J. Ross-Ibarra, and J. Doebley. 2011. Identification of a functional transposon insertion in the maize domestication gene tb1. Nature Genetics 43:1160-1163.

Sumiyama, K., and N. Saitou. 2011. Loss-of-Function Mutation in a Repressor Module of Human-Specifically Activated Enhancer HACNS1. Molecular Biology and Evolution 28:3005-3007.

Swanson, C. I., N. C. Evans, and S. Barolo. 2010. Structural rules and complex regulatory circuitry constrain expression of a Notch- and EGFR-regulated eye enhancer. Developmental Cell 18:359-370.

Swanson, C. I., D. B. Schwimmer, and S. Barolo. 2011. Rapid evolutionary rewiring of a structurally constrained eye enhancer. Current Biology 21:1186-1196.

Team, R. D. C. 2005. A language and environment for statistical computing. . R Foundation for Statistical Computing, Vienna, Austria.

The, I., G. Hannigan, G. Cowley, S. Reginald, Y. Zhong, J. Gusella, I. Hariharan, and A. Bernards. 1997. Rescue of a Drosophila NF1 Mutant Phenotype by Protein Kinase A. Science 276:791-794.

Tirosh, I., A. Weinberger, D. Bezalel, M. Kaganovich, and N. Barkai. 2008. On the relation between promoter divergence and gene expression evolution. Molecular Systems Biology 4:159.

Tong, J. J., S. E. Schriner, D. McCleary, B. J. Day, and D. C. Wallace. 2007. Life extension through neurofibromin mitochondrial regulation and antioxidant therapy for neurofibromatosis-1 in Drosophila melanogaster. Nature Genetics 39:476-485.

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

Tsankov, A. M., D. A. Thompson, A. Socha, A. Regev, and O. J. Rando. 2010. The Role of Nucleosome Positioning in the Evolution of Gene Regulation. PLoS Biology 8:e1000414.

Uemura, O., Y. Okada, H. Ando, M. Guedj, S. Higashijima, T. Shimazaki, N. Chino, H. Okano, and H. Okamoto. 2005. Comparative functional genomics revealed conservation and diversification of three enhancers of the isl1 gene for motor and sensory neuron-specific expression. Dev Biol 278:587-606.

Vandaele, C., M. Coulon-Bublex, P. Couble, and B. Durand. 2001. Drosophila regulatory factor X is an embryonic type I sensory neuron marker also expressed in spermatids and in the brain of Drosophila. Mechanisms of development 103:159-162.

Visel, A., E. M. Rubin, and L. A. Pennacchio. 2009. Genomic views of distant-acting enhancers. Nature 461:199-205.

Walter, M. F., B. C. Black, G. Afshar, A.-Y. Kermabon, T. R. F. Wright, and H. Biessmann. 1991. Temporal and spatial expression of the yellow gene in correlation with cuticle formation and DOPA decarboxylase activity in drosophila development. Developmental Biology 147:32-45.

Weirauch, M. T., and T. R. Hughes. 2010. The Evolution of Combinatorial Gene Regulation in Fungi. Trends in Genetics 26:66-74.

Weiszmann, R., A. S. Hammonds, and S. E. Celniker. 2009. Determination of gene expression patterns using high-throughput RNA in situ hybridization to whole-mount Drosophila embryos. Nature Protocols 4:605-618.

Wergedal, J. E., K. Veskovic, M. Hellan, C. Nyght, W. Balemans, C. Libanati, F. M. Vanhoenacker, J. Tan, D. J. Baylink, and W. Van Hul. 2003. Patients with Van Buchem Disease, an Osteosclerotic Genetic Disease, Have Elevated Bone Formation Markers, Higher Bone Density, and Greater Derived Polar Moment of Inertia than Normal. Journal of Clinical Endocrinology & Metabolism 88:5778-5783.

Werner, T., S. Koshikawa, T. M. Williams, and S. B. Carroll. 2010a. Generation of a novel wing colour pattern by the Wingless morphogen. Nature 464:1143-1148.

Werner, T., S. Koshikawa, T. M. Williams, and S. B. Carroll. 2010b. Generation of a novel wing colour pattern by the Wingless morphogen. Nature 464:1143-1148.

Werren, J. H., S. Richards, C. A. Desjardins, O. Niehuis, J. r. Gadau, J. K. Colbourne, and T. N. G. W. Group. 2010. Functional and Evolutionary Insights from the Genomes of Three Parasitoid Nasonia Species. Science 327:343-348.

West, A. G., M. Gaszner, and G. Felsenfeld. 2002. Insulators: many functions, many mechanisms. Genes &amp; Development 16:271-288.

Wilczynski, B., and E. E. M. Furlong. 2010. Challenges for modeling global gene regulatory networks during development: Insights from Drosophila. Developmental Biology 340:161-169.

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

Wilson, M. D., N. L. Barbosa-Morais, D. Schmidt, C. M. Conboy, L. Vanes, V. L. J. Tybulewicz, E. M. C. Fisher, S. Tavare, and D. T. Odom. 2008. Species-Specific Transcription in Mice Carrying Human Chromosome 21. Science 322:434-438.

Wittkopp, P. J. 2006. Evolution of cis-regulatory sequence and function in Diptera. Heredity 97:139-147.

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

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

Wittkopp, P. J., and G. Kalay. 2012. Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. Nat Rev Genet 13:59-69.

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

Wittkopp, P. J., J. R. True, and S. B. Carroll. 2002a. Reciprocal functions of the Drosophila Yellow and Ebony proteins in the development and evolution of pigment patterns. Development 129:1849-1858.

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

Wittkopp, P. J., K. Vaccaro, and S. B. Carroll. 2002c. Evolution of yellow Gene Regulation and Pigmentation in Drosophila. Current Biology 12:1547-1556.

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

Woolfe, A., and G. Elgar. 2008. Organization of conserved elements near key developmental regulators in vertebrate genomes. Advances in Genetics 61:307-338.

Woolfe, A., M. Goodson, D. K. Goode, P. Snell, G. K. McEwen, T. Vavouri, S. F. Smith, P. North, H. Callaway, K. Kelly, K. Walter, I. Abnizova, W. Gilks, Y. J. K. Edwards, J. E. Cooke, and G. Elgar. 2005. Highly Conserved Non-Coding Sequences Are Associated with Vertebrate Development. PLoS Biology 3:e7.

Wratten, N., A. McGregor, P. Shaw, and G. Dover. 2006. Evolutionary and functional analysis of the tailless enhancer in Musca domestica and Drosophila melanogaster. Evolution & Development 8:6-15.

Xia, A.-H., Q.-X. Zhou, L.-L. Yu, W.-G. Li, Y.-Z. Yi, Y.-Z. Zhang, and Z.-F. Zhang. 2006. Identification and analysis of YELLOW protein family genes in the silkworm, Bombyx mori. BMC Genomics 7:195.

Yeates, D. K., and B. M. Wiegmann. 1999. Congruence and Controversy: Toward a Higher-Level Phylogeny of Diptera. Annual Review of Entomology 44:397-428.

Yuh, C. H., and E. H. Davidson. 1996. Modular cis-regulatory organization of Endo16, a gut-specific gene of the sea urchin embryo. Development 122:1069-1082.

Yuh, C. H., J. G. Moore, and E. H. Davidson. 1996. Quantitative functional interrelations within the cis-regulatory system of the S. purpuratus Endo16 gene. Development 122:4045-4056.

CHAPTER 2

Nomadic enhancers: Tissue-specific *cis*-regulatory elements of *yellow* have divergent genomic positions among Drosophila species.

Abstract

cis-regulatory DNA sequences known as enhancers control gene expression in space and time. They are central to metazoan development and are often responsible for changes in gene regulation that contribute to phenotypic evolution. Here, we examine the sequence, function, and genomic location of enhancers controlling tissue- and cell-type specific expression of the *yellow* gene in six Drosophila species. *yellow* is required for the production of dark pigment and its expression has evolved largely in concert with divergent pigment patterns. Using *Drosophila melanogaster* as a transgenic host, we examined the expression of reporter genes in which either 5' intergenic or intronic sequences of *yellow* from each species controlled the expression of Green Fluorescent Protein. Surprisingly, we found that sequences controlling expression in the wing veins, as well as sequences controlling expression in epidermal cells of the abdomen, thorax, and wing, were located in different genomic regions in different species. By contrast, sequences controlling expression in bristle-associated cells were located in the intron of all species. Differences in the precise pattern of spatial

expression within the developing epidermis of *D. melanogaster* transformants usually correlated with adult pigmentation in the species from which the *cis*-regulatory sequences were derived, which is consistent with *cis*-regulatory evolution affecting *yellow* expression playing a central role in Drosophila pigmentation divergence. Sequence comparisons among species favored a model in which sequential nucleotide substitutions were responsible for the observed changes in *cis*-regulatory architecture. Taken together, these data demonstrate frequent changes in *yellow cis*-regulatory architecture among Drosophila species. Similar analyses of other genes, combining *in vivo* functional tests of enhancer activity with *in silico* comparative genomics, are needed to determine whether the pattern of regulatory evolution we observed for *yellow* is characteristic of genes with rapidly evolving expression patterns.

The work in Chapter 2 is published with the following citation:

Kalay, G. and P. J. Wittkopp (2010). "Nomadic enhancers: tissue-specific cis-regulatory elements of yellow have divergent genomic positions among Drosophila species." PLoS genetics **6**(11): e1001222.

Author summary

In order for a gene to be active, it must be turned on, or “expressed”. Instructions determining when, where, and how much a gene will be expressed are encoded by DNA sequences known as enhancers. The precise DNA sequence of a particular enhancer changes over evolutionary time, which may or may not change its effects on gene expression. Many genes are controlled by multiple enhancers and prior work has shown that the location of these enhancers within the genome tends to remain stable for long periods of evolutionary time. Here, we examine the enhancers controlling expression of a gene (yellow) involved in generating pigmentation diversity among fruit fly (Drosophila) species. Surprisingly, we find that not only have the sequence and function of individual enhancers changed among Drosophila species, but so has the location of these enhancers within each the genome of each species. This finding is important because it demonstrates a type of evolutionary change affecting DNA sequence elements critical for gene expression that is currently under appreciated and should be considered when searching for enhancers in future studies.

Introduction

The production of a complex, multi-cellular organism requires transcription of a subset of the genome in each cell. This process, known as gene expression, is controlled by *cis*-regulatory DNA sequences that interact with *trans*-regulatory proteins and RNAs. These *cis*-regulatory sequences include “enhancers”, which contain binding sites for transcription factors. The specific combination of transcription factor binding sites within an enhancer determines its activity and specifies the timing, location, and abundance of expression for the gene it regulates. Many genes, especially those involved in development, are controlled by multiple enhancers, each of which controls a subset of the gene’s total expression pattern and can be located 5’, 3’ or in an intron of the gene whose transcription it regulates. Like all DNA, *cis*-regulatory sequences are subject to the unavoidable process of mutation, which – over evolutionary time – can change enhancer sequence, enhancer function, and the genomic location of enhancers relative to the gene whose expression they control.

Comparing the *cis*-regulatory architecture of orthologous genes among species reveals how they evolve as well as which features are essential for their activity. Conserved sequences between orthologous enhancers represent putatively functional elements (e.g., (Langeland and Carroll 1993; Lukowitz et al. 1994)), but conservation of DNA sequence is not strictly required for conservation of enhancer function: transcription factor binding sites are often degenerate and comparable enhancer functions can be produced by multiple arrangements of these sites (Ludwig et al. 1998; Romano and Wray 2003a; Hare et al. 2008b). Compared to enhancer sequence, enhancer location within the genome (relative to exonic sequences of the associated gene) appears to be more constrained. For example, the location of enhancers is conserved for the *even-skipped* gene between Drosophila and Sepsid species (Hare et al. 2008b), which diverged over 100 million years ago, and for six Dorsal target genes between Drosophila and

Anopheles or Tribolium (Cande et al. 2009b), which diverged over 200 million years ago. Indeed, conservation of enhancer location within the genome is something that many researchers rely upon in their search for orthologous enhancers.

Here, we investigate the evolution of *cis*-regulatory architecture controlling expression of the Drosophila *yellow* gene. Yellow is required for the production of dark melanic pigment in insects (Arakane et al. ; Wittkopp et al. 2002b; Futahashi et al. 2008), and its expression during late pupal stages has evolved in a manner that often correlates with the distribution of melanins in adults (Wittkopp et al. 2002d; Gompel et al. 2005b; Jeong et al. 2006b; Prud'homme et al. 2006b). In *D. melanogaster*, *yellow* expression is controlled by multiple tissue-specific enhancers, with enhancers driving expression in the pupal wing, abdomen, and thorax located 5’ of the *yellow* gene and an enhancer driving expression in bristle-associated cells located within its lone intron (Geyer and Corces 1987a; Martin et al. 1989b; Wittkopp et al. 2002d; Jeong et al. 2006b) (Figure 2-1). Comparisons of *yellow* expression and regulation among species suggest that changes in *cis*-regulatory activity are most often responsible for divergent *yellow* expression patterns (Wittkopp et al. 2002d; Gompel et al. 2005b; Jeong et al. 2006b; Prud'homme et al. 2006b; Werner et al. 2010b), although changes in *trans*-regulatory factors also contribute to expression divergence in some species (Wittkopp et al. 2002d; Werner et al. 2010b)). Changes in the spatial pattern of *yellow* expression within the developing abdomen result from changes in orthologous enhancers located in the 5’ intergenic sequences of *yellow* (Wittkopp et al. 2002d; Jeong et al. 2006b), and convergent *yellow* expression in “spots” on the developing wing results from enhancers that evolved in the 5’ intergenic region of one species and in the intron of another (Gompel et al. 2005b; Prud'homme et al. 2006b; Werner et al. 2010b).

To examine the evolution of *yellow cis*-regulatory architecture more comprehensively and systematically, we determined the enhancer activity of sequences 5’ of *yellow* and in its intron for six species spanning the phylogenetic tree of the genus Drosophila. These species include members of both the *Drosophila* (*D. mojavensis, D. virilis,* and *D. grimshawi*) and Sophophora (*D. melanogaster, D. pseudoobscura,* and *D. willistoni)* subgenera and have pairwise divergence times ranging from approximately 20 to 40 million years ago (Russo et al. 1995; Spicer and Bell 2002). Surprisingly, we found that the location of *yellow* enhancer activity controlling expression in a particular tissue- or cell-type differed frequently among species, with only the enhancer controlling bristle-associated expression located in the same genomic region of all species. These differences in *cis*-regulatory architecture were accompanied by differences in enhancer activity that often correlated with species-specific pigment patterns, as expected based on prior studies (Wittkopp et al. 2002d; Gompel et al. 2005b; Jeong et al. 2006b; Prud'homme et al. 2006b; Werner et al. 2010b). Sequence comparisons between pairs of species showed no clear evidence of duplications or transpositions near *yellow*, suggesting that differences in enhancer location among species evolved by sequential sequence substitutions, one or a few nucleotides at a time. To the best of our knowledge, such extensive and rapid turnover in the genomic location of enhancers has not been observed for any other eukaryotic gene.

Results

To determine the *cis*-regulatory architecture of *yellow* in each of six *Drosophila* species, we constructed reporter genes that used species-specific 5’ intergenic or intronic sequences of *yellow* to drive expression of a nuclear Green Fluorescent Protein (GFP) in transgenic *D. melanogaster*. The 5’ intergenic regions surveyed began near a highly-conserved region of sequence (Figure 2-6) located 5’ of the previously characterized wing and body enhancers of *D. melanogaster yellow* (Geyer and Corces 1987a; Martin et al. 1989b; Wittkopp et al. 2002d; Jeong et al. 2006b; Werner et al. 2010b) and extended 3’ to the beginning of the first exon of *yellow* (Figure 2-2). This region includes all of the 5’ intergenic DNA contained within *yellow* transgenes that fully rescue *yellow* null mutant phenotypes in *D. melanogaster* (Geyer and Corces 1987a) and *D. virilis* (Wittkopp et al. 2002d), suggesting that these constructs are likely to contain all 5’ enhancers affecting *yellow* expression. The intronic constructs began and ended with sequences in the first and second exons, respectively. DNA fragments tested ranged from 4 to 9.8 kb for the 5’ intergenic regions and from 2.7 to 6.7 kb for the intronic regions (Figure 2-2). Each of the twelve reporter genes was independently integrated into the same pre-determined location of the *D. melanogaster* genome using the phiC31 integrase system (Groth et al. 2004), and expression of the GFP reporter gene in transgenic pupae 70-80 hours after puparium formation was examined by confocal microscopy.

[Note: Reporter genes containing 5’ intergenic and intronic sequences from *D. subobscura* were also constructed and analyzed; however, because the 5’ intergenic region surveyed in *D. subobscura* did not extend to the highly-conserved region, these data are presented and discussed only in Figure 2-7 and its associated legend.]

Genomic location of tissue-specific enhancers differs among species

All DNA fragments tested were sufficient to activate GFP expression in at least one tissue during the pupal stage examined (Figure 2-3). Reporter genes containing 5’ intergenic and intronic sequences from *D. melanogaster* drove expression patterns consistent with prior studies (Geyer and Corces 1987a; Martin et al. 1989b; Wittkopp et al. 2002d; Jeong et al. 2006b; Werner et al. 2010b): the 5’ intergenic sequence drove expression in the epidermal cells of the abdomen, thorax and wing (Figure 2-3B), whereas the intronic sequence drove expression in bristle-associated cells (Figure 2-3C). We also observed faint expression in wing veins activated by the *D. melanogaster* intronic sequence (Figure 2-3C, arrows) -- an enhancer activity that (to the best of our knowledge) has not previously been reported in *D. melanogaster*. Reporter gene expression was similarly used to infer the location of tissue- and cell-type specific enhancers in each of the other five species. Locations for enhancers that drive expression in the epidermal cells of the abdomen, thorax, wing, and head; in the wing veins; and in bristle-associated cells are summarized in the following paragraphs.

For each species, enhancers driving expression in epidermal cells of the abdomen, thorax, wing, and (when expression was present) head were typically found in the same genomic region; however, the location of this region differed among species and half of the species showed evidence of epidermal cell enhancers in both the 5' intergenic and intronic regions. Enhancers driving expression in epidermal cells of the abdomen, thorax, and wing were observed in the 5' intergenic regions of all three Sophophora species (i.e., *D. melanogaster*, *D. pseudoobscura*, and *D. willistoni*) and *D. virilis* from the Drosophila subgenus (Figures 2-3B, E, H, N) as well as in the introns of *D. pseudoobscura* and all three species from the Drosophila subgenus (i.e., *D. mojavensis, D. virilis,* and *D. grimshawi*) (Figures 2-3F, L, O and R). In addition, the intron from *D. willistoni* drove expression in the epidermal cells of the thorax and wing (Figure 2-3I), and the *D. grimshawi* 5' intergenic region drove expression in a small region of epidermal cells flanking two of the wing veins (Figure 2-3Q, arrows). Expression in head epidermal cells was observed only in *D. pseudoobscura* and *D. virilis*, with the enhancer controlling this expression located in the 5' intergenic or intronic regions of these species, respectively (Figures 2-3E and O).

The genomic location of enhancers driving expression in wing veins was also variable among species. In the subgenus Sophophora, the two most closely related species, *D. melanogaster* and *D. pseudoobscura*, both showed this enhancer activity in the intron (Figures 2-3C and F, arrows), whereas the more distantly related *D. willistoni* showed wing vein enhancer activity in the 5' intergenic sequence (Figure 2-3H, arrow). In the subgenus Drosophila, both 5' intergenic and intronic sequences from *D. mojavensis* and *D. virilis* drove expression in the wing veins (Figures 2-3K, L, N, and O, arrows), but no wing vein expression was observed from either reporter gene containing *D. grimshawi* sequence (Figures 2-3Q and R).

Expression in bristle-associated cells of both the body and wing was controlled by intronic sequences from all six species, making it the only *yellow* enhancer activity whose genomic location appears to be conserved within the genus Drosophila (Figure 2-3C, F, I, L, O, and R).

Divergent activity of *yellow* enhancers often correlates with divergent pigmentation

The spatial patterns of reporter gene expression in epidermal cells of the abdomen, thorax, and (less frequently) wing often differed between species (Figure 2-3). With few exceptions (noted below), sequences from each species activated GFP expression in transgenic *D. melanogaster* hosts in patterns that correlated with adult pigmentation of the species from which the enhancer sequences were derived. In the abdomen, for example, *D. melanogaster, D. willistoni,* and *D. grimshawi* all have dark stripes at the posterior edge of each dorsal abdominal segment (Figures 2-3A, G, and P) and show similar stripes of reporter gene expression in each abdominal segment driven by either their 5' intergenic or intronic sequences (Figures 2-3B, H, and R). *D. mojavensis*, however, also has pigment stripes on its dorsal abdomen, but the weak abdominal reporter gene expression observed was not restricted to these stripes (Figure 2-3L). In addition, *D. mojavensis* has a series of pigment spots on its head and thorax (Figure 2-3J), and *D. grimshawi* has dark pigments along the dorsal midline in the abdomen and in the thorax (Figure 2-3P), neither of which are reflected in the expression patterns of the corresponding species-specific reporter genes (Figures 2-3K, L, Q, and R). Finally,*D. pseudoobscura* and *D. virilis* have an overall dark body color and faint stripes on the thorax (Figures

2-3D and M), all of which are reflected in the reporter gene expression patterns for both species (Figures 2-3E, F, N, and O).

Partial correlations between reporter gene expression and adult pigmentation were also seen in the wing. *D. virilis* has a visible spot of dark pigment surrounding one of its cross-veins (Figure 2-3M), and *D. grimshawi* has an elaborate pattern of pigment spots (Figure 2-3P). The 5' intergenic region from *D. virilis* drove higher levels of expression in cells that will give rise to the pigmented spot surrounding L4-L5 cross-vein than in the rest of the wing (Figure 2-3N, arrowhead), whereas the *D. grimshawi* intron drove elevated expression in a subset of wing epidermal cells in a pattern that did not correlate well with adult *D. grimshawi* wing pigmentation (Figure 2-3R). Interestingly, the *D. pseudoobscura* intron drove elevated expression in an anterior spot of the wing (Figure 2-3F, arrowhead) despite the fact that *D. pseudoobscura* lacks any obvious dark pigment patterns in this region.

Nomadic enhancers: moving existing elements or *de novo* construction and destruction?

As described above, similar tissue-specific enhancer activities were found in different genomic regions among the species surveyed. Such changes in *cis*-regulatory architecture can be achieved through (1) the movement of existing enhancers via duplications and/or transpositions of DNA sequence or (2) the *de novo* construction or destruction of transcription factor binding sites individually via sequential nucleotide substitutions. Each of these mechanisms is expected to produce a different pattern of sequence similarity between species. For example, consider *D. melanogaster*, which has an enhancer driving expression in abdominal epidermal cells in its 5' intergenic region (Figure 2-3B), and *D. pseudoobscura*, which has two enhancers driving expression in abdominal epidermal cells located in its 5' intergenic and intronic regions (Figures 2-3E and F). If the intronic enhancer in *D. pseudoobscura* resulted from a duplication of the

5’ enhancer shared with *D. melanogaster*, sequence similarity is expected between the 5’ region of *D. melanogaster* and the intron of *D. pseudoobscura* as well as between the 5’ intergenic and intronic sequences of *D. pseudoobscura* itself. If, however, a more gradual sequence substitution process caused either the loss of abdominal epidermal cell enhancer activity in the *D. melanogaster* intron or the gain of this activity in the *D. pseudoobscura* intron, regions of sequence similarity are expected to be collinear between species. That is, the introns of both species should share greater sequence similarity with each other than either does with the other species’ 5’ intergenic sequence and vice versa.

To try to distinguish between these mechanisms, we performed pairwise comparisons of *yellow* genes and their 5’ intergenic sequences for all six species. As expected, significant sequence similarity was observed between homologous exons for all pairs of species (Figure 2-4). Outside of these regions, very little sequence similarity was observed for all but the most closely related pairs of species in each subgenus: *D. melanogaster* and *D. pseudoobscura* in the Sophophora subgenus, and *D. mojavensis* and *D. virilis* in the Drosophila subgenus. These two pairs of species provide the most power for investigating the molecular mechanisms responsible for interspecific differences in enhancer location. In both cases, one species in the pair has enhancer activity driving epidermal cell expression in the abdomen, thorax, and wing only in the 5’ intergenic region or only in the intron, whereas the other member of the pair has similar activities in both the 5’ intergenic region and the intron. Despite these differences in the genomic location of enhancers with similar tissue-specificity, we observed only collinear regions of sequence similarity (Figure 2-4, red and blue arrows). Such a pattern favors a model in which enhancers have been gained or lost through sequential sequence substitutions.

Discussion

We found that the *cis*-regulatory architecture of *yellow* has changed repeatedly during the ~40 million years since the six Drosophila species we examined last shared a common ancestor. This includes changes in the activity of homologous tissue-specific enhancers as well as changes in their relative genomic location. Sequence comparisons between the most closely related species examined showed no evidence of duplications or transpositions, suggesting that this diversity may have arisen through the gradual accumulation of sequence differences one (or a few) nucleotides at a time. As discussed below, these data provide insight into the independence of tissue-specific enhancers, the relationship between *yellow* enhancers and pigmentation divergence, and the evolution of *cis*-regulatory architecture.

Evolutionary constraint suggests interactions between tissue-specific enhancers

Comparative studies that examine *cis*-regulatory sequences in an evolutionary context can uncover features overlooked by dissecting *cis*-regulatory sequences from a single species. For example, studies of *D. melanogaster yellow* identified non-overlapping DNA sequences that are necessary and sufficient to activate expression in epidermal cells of the body (i.e., abdomen and thorax) or wing, suggesting the presence of two distinct tissue-specific enhancers (Geyer and Corces 1987a; Wittkopp et al. 2002d). We found that these “wing” and “body” enhancer activities colocalize to the same genomic region in most species despite frequent evolutionary changes in the relative position of this region (Figure 2-5). This suggests that these enhancers are not fully independent, but rather interact in a way that constrains their evolution. For example, they might require close proximity to function properly at the native *yellow* locus because they share transcription factor binding sites and/or chromatin structure that promotes expression in pupal epidermal cells. Such colocalization was not observed for enhancers driving expression in bristle-associated cells or wing veins. Therefore, we propose that three evolutionarily independent enhancer modules regulate *yellow* expression: one controlling expression in bristle-associated cells, one controlling expression in the wing veins, and one controlling expression in the epidermal cells of the abdomen, thorax, head, and/or wing. Consistent with this proposal, a DNA fragment containing both the previously defined “body” and “wing” enhancers drives reporter gene expression in epidermal cells of the abdomen that is more representative of endogenous *D. melanogaster yellow* expression in those cells than that driven by a fragment containing the “body” enhancer alone (Jeong et al. 2006b).

yellow **enhancer activity often, but not always, evolves with pigmentation**

In Drosophila, *cis*-regulatory changes affecting *yellow* expression often correlate with changes in pigmentation among species (Wittkopp et al. 2002d; Gompel et al. 2005b; Jeong et al. 2006b; Prud'homme et al. 2006b; Werner et al. 2010b), suggesting that they have contributed to the evolution of this trait. Indeed, we observed a correlation between pigmentation and enhancer activity in most of our dataset; however, not all pigment patterns were reflected in reporter gene expression. For example, pigment spots on the body of *D. mojavensis* and on the wings of *D. grimshawi* were not observed in the expression pattern of either of the reporter genes from these species. This could be because these particular pigment patterns are controlled by another pigmentation gene such as *tan* (True et al. 2005; Jeong et al. 2008; Wittkopp et al. 2009). Alternatively, enhancers driving *yellow* expression in these patterns may be located outside of the regions surveyed; an additional wing enhancer was found in *D. grimshawi* 5’ of intergenic region we tested (T. Werner and S.B. Carroll, personal communication). Finally, *trans*-acting factors controlling *yellow* expression may have diverged between *D. melanogaster* and *D. mojavensis* or *D. grimshawi* such that sequences drive expression in a different pattern when inserted into the *D. melanogaster* genome than they do in their native species. Such *trans*-regulatory changes are known to exist between *D. melanogaster* and *D. virilis* (Wittkopp et al. 2002d) and between *D. melanogaster* and *D. guttifera* (Werner et al. 2010b).

In addition to pigment patterns not reflected in reporter gene expression, we also observed reporter gene expression not reflected in pigment patterns. Intronic sequences from *D. pseudoobscura* activated reporter gene expression in an anterior region of *D. melanogaster* wings despite the fact that adult *D. pseudoobscura* lack pigmentation in this area (Figure 2-3F, arrowhead). This expression pattern does not appear to be an artifact of the heterologous transgenic host because a similar pattern is seen in native *D. pseudoobscura* Yellow expression (see Figure 1 in (Gompel et al. 2005b)). Interestingly, *D. tristis*, which is a member of the obscura group to which *D. pseudoobscura* also belongs, has a similar pattern of *yellow* expression in pupal wings controlled by an intronic enhancer and does display a corresponding spot of pigmentation on its adult wings (Prud'homme et al. 2006b). This spot of wing pigmentation appears to be a derived trait in the obscura group (Prud'homme et al. 2006b), thus the presence of this expression pattern in *D. pseudoobscura* suggests that the novel *yellow* enhancer activity in this wing spot preceded other changes, such as a coincident decrease in Ebony protein expression (Wittkopp et al. 2002b; Gompel et al. 2005b), that are also required for wing spot formation.

Evolutionary changes responsible for the dynamic *cis*-regulatory architecture of *yellow*

Examining divergent phenotypes in concert with a phylogenetic tree allows inferences to be made about the evolutionary changes that led to the observed trait diversity. To this end, Figure 2-5 shows the phylogenetic relationships among the species surveyed alongside a summary of the genomic locations of *yellow* enhancers from each species. Enhancer activity was considered present if reporter gene expression was observed in the tissue- or cell-type indicated regardless of the precise spatial pattern within that tissue.

To determine the evolutionary changes that gave rise to the observed diversity of *cis*-regulatory architecture, we must first infer the genomic locations of enhancers in the common ancestor of the six species studied. To do this, we considered each enhancer activity independently. The historical genomic location of bristle enhancer activity could be inferred with the most confidence: all six species showed bristle enhancer activity only in the intron, strongly suggesting that the common ancestor of these six species also had a bristle enhancer in this region. The ancestral locations of the wing vein and epidermal cell enhancers is less clear; these enhancer activities were found in the 5’ intergenic region, in the intron, and in both of these regions depending on the species surveyed.

Inferring the most likely genomic location(s) of wing vein and epidermal cell enhancers in the common ancestor requires an assumption about the relative likelihood of enhancer gain and enhancer loss in different lineages. Because mutations are expected to disrupt transcription factor binding sites more often than they are expected to create new ones, we assume that the loss of enhancer activity is more likely in all lineages than the gain of a novel tissue-specific enhancer. On the basis of this assumption, the most parsimonious explanation for the observed data is that the common ancestor had enhancers in both the 5’ intergenic and intronic regions of *yellow* that drove expression in the wing veins as well as in the abdomen, thorax, and wing epidermal cells. Such a scenario involves at least one loss of enhancer activity in the lineage leading to each of the species surveyed except *D. virilis*, as shown in Figure 2-5. While we find a common ancestor with redundant enhancers in the 5’ intergenic and intronic regions for both the wing veins and epidermal cells surprising, overlapping enhancers with similar tissue- and cell-type specific activities have been identified for other genes (e.g., (Helms et al. 2000; Pappu et al. 2005; Uemura et al. 2005; Jeong et al. 2006c; Cretekos et al. 2008; Hong et al. 2008; Frankel et al. 2010b)) and scenarios involving a common ancestor with wing vein and/or epidermal cell enhancer activity in only one genomic region include multiple gains and losses in most lineages, which is presumably even less likely.

Regardless of the specific gains, losses, and/or relocations of *yellow* enhancers that occurred over the last 40 million years, it is clear that the genomic location of enhancer activities within and surrounding the *yellow* gene has changed multiple times. This finding is contrary to recent studies of seven other genes showing conserved genomic locations of Drosophila enhancers in species that diverged over 100 million years ago (e.g., (Hare et al. 2008b; Cande et al. 2009b)) , and challenges the assumption of conserved enhancer location that often underlies searches for orthologous enhancers. At least one other Drosophila gene (i.e., *twist*) has analogous differences in enhancer location between species (Pan et al. 1994a); however, the frequency of such changes on a genomic scale remains unknown. Given the rapid sequence divergence of even functionally conserved enhancers (reviewed in (Wittkopp 2006)), changes in enhancer location are unlikely to be detected by sequence alignments alone, underscoring the importance of supplementing *in silico* comparative genomics with *in vivo* functional tests.

Materials and Methods:

Isolating yellow *BAC clones*

For five of the six species used in this study (*D. pseudoobscura*, *D. willistoni*, *D. mojavensis*, *D. virilis*, and *D. grimshawi*), BAC libraries (CHORI-222, DW_Ba, DM_CBa, DV_VBa and DG_Ba, respectively) were screened for clones containing *yellow* as well as its flanking genes. Nylon filters containing arrayed clones from the BAC libraries were obtained from BACPAC Resources (CHORI-222) and Arizona Genomics Institute (AGI) (DW_Ba, DM_CBa, DV_VBa and DG_Ba), and screened with [alpha-32-P]-labeled, random hexamer-primed probes synthesized using PCR amplicons from exons of the *yellow* gene; the *CG3777* gene, which is located 5' of *yellow*; and either the *CG4165* (*D. mojavensis*) or *achete* (all other species) gene, both of which are located 3' of *yellow*. (Primers and PCR conditions used to amplify the DNA template for each probe are available upon request.) Probe synthesis was performed as described in Molecular Cloning (Sambrook and Russell 2001). Unincorporated radionucleotides were removed using CentriSpin columns (Princeton Separations). Purified radioactive probes were denatured at 100^0C for 5 minutes and placed on ice until they were added to the hybridization buffer containing the appropriate species specific BAC filter. BAC filter screening conditions and buffer recipes were as described in the AGI BAC Filter Manual available from the Arizona Genomics Institute (http://www2.genome.arizona.edu/research/protocols_bacmanual). After hybridizing each filter with a radioactive probe, the filter was washed and exposed to Kodak BioMax XAR films for 72 hours @ -80^0C and developed.

Radiographs were used to identify clones as directed by the filter manufacturers (Arizona Genomics Institute and BACPAC Resources), and BACs that hybridized to all three probes were ordered. Upon receipt, each BAC clone was tested for the presence of *CG3777*, *yellow*, and *achete* or *CG4165* using PCR amplification. Table 2-1 lists all BAC clones found to contain *yellow* and at least one flanking gene. For *D. willistoni*, *D. mojavensis*, *D. virilis*, and *D. grimshawi*, BAC clones with code numbers 10L5, 4J24, 1A7 and 23K7, respectively, were used for reporter gene construction. For *D. melanogaster*, the RP98-13J2 BAC clone from the Roswell Park Cancer Institute Drosophila BAC Library, which was identified computationally and confirmed by PCR to contain *CG3777*, *yellow* and *achete,* was used for reporter gene construction. Note that none of the *D. pseudoobscura* BAC clones containing *yellow* had sufficient 5' sequence to be used for reporter gene construction.

Constructing reporter genes

For each species, 5’ intergenic and intronic regions of *yellow* were cloned into a plasmid containing piggyBac transposable element arms, a 3xP3-Enhanced Green Fluorescent Protein (EGFP) marker driving cytoplasmic GFP expression in the eyes (Horn and Wimmer 2000), and a 300 bp attB site (Groth et al. 2004; Bischof et al. 2007) that we amplified from the pTA-attB plasmid provided by Michele Calos (Stanford University) and inserted into the unique XbaI site. As described in the main text, the 5’ end of the 5’ intergenic sequences was defined by the highly conserved region shown in Figure 2-6. The 5’ intergenic and intronic sequences from *D. melanogaster, D. subobscura*, *D. pseudoobcsura,* and the intron of *D. virilis yellow* were PCR amplified from BAC RP98-13J2, plasmid ysub-pBac (Wittkopp et al. 2002d), genomic DNA extracted from *D. pseudoobscura* (UCSD stock number 14011-0121.94), and plasmid yvir-pBac (Wittkopp et al. 2002d), respectively. Primer sequences used for these amplifications are available upon request. PCR products were ligated to the PCR 2.1 TOPO vector (Invitrogen), fully sequenced to identify clones with no PCR introduced mutations, and subcloned into the piggyBac-EGFP vector described above using the unique AscI restriction site.

For *D willistoni, D. mojavensis, D. grimshawi,* and the *D. virilis* both the 5’ intergenic and intronic regions were cloned into the piggyBac-EGFP vector using recombineering (http://recombineering.ncifcrf.gov/). Briefly, PCR was used to amplify 450-500 bp homology arms corresponding to the 5’ (left arm) and the 3’ (right arm) end of each target DNA sequence. PCR sewing was used to combine the left and right arms into a single fragment with a unique NheI restriction site between them. These DNA fragments were subcloned into PCR 2.1 TOPO, fully sequenced to identify clones without PCR introduced mutations, and subcloned into the piggyBac-EGFP vector using the unique AscI restriction site. Each piggyBac vector containing a species-specific pair of homology arms was linearized using the introduced NheI restriction site and electroporated into SW102 cells containing the *yellow* BAC from the appropriate species.

Electroporation was conducted using Eppendorf Electroporator 2510 at 1250 Volts, with time constants ranging between 4.5-5. Following electroporation, SW102 cells were incubated in 1 ml LB at 30^0C rotator for 1-1.5 hours, spread on LB agar plates supplemented with ampicillin (50 ug/ml), and grown overnight at 30^0C to select for cells containing a circularized piggyBac-EGFP plasmid harboring the DNA of interest. Primers located in the piggyBac vector and in the target DNA sequences were paired to screen colonies for the existence and the direction of the DNA region of interest using PCR. Positive clones were confirmed by diagnostic digests using restriction enzymes specifically chosen for each construct, and the inserted DNA was completely sequenced to confirm once again that no experimentally introduced mutations were present. Next, a DNA fragment derived from pSLfa1180fa-nEGFP (Ernst Wimmer, Georg August University, Göttingen) containing an hsp70 promoter and the coding sequence for a nuclear EGFP protein was cloned into each piggyBac plasmid using the unique FseI restriction site. The resulting DNA transgene constructs were confirmed using appropriate diagnostic digests with restriction enzymes and sent to Genetics Services, Inc. (Cambridge, MA) where they were injected into the *w*$^-$; attP-40 line of *D. melanogaster* (Markstein et al. 2008). This line contains a transgene expressing the φC31 site-specific integrase enzyme (Bischof et al. 2007), which causes the targeted integration of each attB-containing piggyBac construct into the attP site on the *D. melanogaster* 2nd chromosome. An “empty” piggyBac plasmid lacking any *yellow* sequence was also transformed into *D. melanogaster* and analyzed as a control to determine background levels of GFP expression.

Analysis of reporter gene expression patterns

Homozygous transgenic *D. melanogaster* lines were obtained by crossing each transgenic *D. melanogaster* genotype to a 2nd chromosome balancer line (*w[*]; Kr [If-1]/CyO; D[1]/TM6B, Tb[+]*; Bloomington stock number 7197), intercrossing the F_1 offspring, and then intercrossing selected homozygous F_2 individuals. Homozygous transgenic animals were imaged at 70-80 hours APF, a stage which is recognized by pigmented wings as well as the presence of visible malpigian tubes on the anterior sides of the abdomen. The pupal case was removed prior to imaging using a probe and a pair of fine forceps.

To prepare the pupal bodies for confocal microscopy, the transparent pupal cuticle was kept in place without any tears and the pupa was mounted on a microscope slide with a drop of water and a coverslip. To prepare the pupal wings for confocal microscopy, the transparent pupal cuticle was removed and the whole fly was submerged in Milli-Q water. After the wings had unfolded, which took about one minute, they were carefully detached from the rest of the pupa at the base of the wing where it connects to the thorax. Using a wide mouth pipette tip, each wing was transferred onto a microscope slide with a drop of water. A coverslip was applied and pressed gently to achieve full expansion of the wings. All specimens were imaged immediately after mounting using a Leica SP5 confocal microscope. Identical settings (e.g., laser power, pinhole size, etc) were used on the confocal microscope for all samples, and all raw confocal images of the same tissue (e.g., wings or bodies) were processed identically in Photoshop.

Sequence analysis

yellow sequences and 5' intergenic DNA from all species except *D. willistoni* were downloaded using the UCSC Genome Browser (Rhead et al.). Specific assemblies and coordinates for each species were as follows: *D. melanogaster*, Apr. 2006 (BDGP R5/dm3) Assembly, chrX:246,727-255,037; *D. pseudoobscura*, FlyBase release r2.11, chrXL_group1e:4227884-4238281; *D. willistoni*, FlyBase release r1.3 scf2_1100000004909:5315142-5325379; D. mojavensis, Aug. 2005 (Agencourt prelim/droMoj2) Assembly, scaffold_6359:2,460,150-2,478,221; D. virilis, Aug 2005 (Agencourt prelim/droVir2) Assembly, scaffold_13042:3,903,783-3,920,981; D. grimshawi, Aug 2005 (Agencourt prelim/droGri1) Assembly, scaffold_24821:2,532,826-2,547,390. Homologous *D. willistoni* sequences were identified and downloaded using the BLAST implementation on FlyBase. These sequences were subject to repeat masking prior to analysis.

Alignments were performed using LASTZ (Release 1.02.00, built January 12, 2010), which was downloaded from Webb Miller's laboratory website (http://www.bx.psu.edu/). This unpublished software replaces the BLASTZ program developed by the same group (Schwartz et al. 2003). Default settings were used except for the ”--mismatch=2,23” option that sets an alternative threshold for the gap-free extension step. The basic structure of this analysis is as follows: all sequences 19 nucleotides long with matches in 12 specific positions were identified as “seeds”; seeds were extended in both directions without gaps until two mismatches were found in each end; extended seeds at least 23 nucleotides long were treated as “high scoring segment pairs” (HSPs); HSPs were converted into anchor points; anchor points were extended in both directions using gapped local alignments; and the coordinates of local alignments output by LASTZ were plotted using R statistical software (Team 2005). The decision to allow a maximum of two mismatches during the gap-free extension stage was arbitrary, whereas the minimum length of extended seeds treated as HSPs (i.e., 23 nucleotides) was determined empirically by randomizing concatenated multi-species *yellow* sequences with the “Shuffle DNA” tool in the web-based “Sequence Manipulation Suite” (Stothard 2000) and iteratively testing length thresholds to find the smallest value that failed to identify any stretches of significant sequence similarity in the randomized sequence. Figure 2-8 shows the result of the same analysis with a decreased length threshold (”--mismatch-2,19”); 40 regions of significant sequence similarity were identified between the real and randomized sequences using these parameters.

Supporting Text

yellow enhancer activity often, but not always, evolves with pigmentation

In *Drosophila*, *cis*-regulatory changes affecting *yellow* expression often correlate with changes in pigmentation among species (Wittkopp et al. 2002b; Gompel et al. 2005; Jeong et al. 2006; Prud'homme et al. 2006; Werner et al. 2010), suggesting that they have contributed to the evolution of this trait. Indeed, we observed a correlation between pigmentation and enhancer activity in most of our dataset; however, not all pigment patterns were reflected in reporter gene expression. For example, pigment spots on the body of *D. mojavensis* and on the wings of *D. grimshawi* were not observed in the expression pattern of either of the reporter genes from these species. This could be because these particular pigment patterns are controlled by another pigmentation gene such as *tan* (True et al. 2005; Jeong et al. 2008; Wittkopp et al. 2009). Alternatively, enhancers driving *yellow* expression in these patterns may be located outside of the regions surveyed; an additional wing enhancer was found in *D. grimshawi* 5' of intergenic region we tested (T. Werner and S.B. Carroll, personal communication). Finally, *trans*-acting factors controlling *yellow* expression may have diverged between *D. melanogaster* and *D. mojavensis* or *D. grimshawi* such that sequences drive expression in a different pattern when inserted into the *D. melanogaster* genome than they do in their native species. Such *trans*-regulatory changes are known to exist between *D. melanogaster* and *D. virilis* (Wittkopp et al. 2002b) and between *D. melanogaster* and *D. guttifera* (Werner et al. 2010).

In addition to pigment patterns not reflected in reporter gene expression, we also observed reporter gene expression not reflected in pigment patterns. Intronic sequences from *D. pseudoobscura* activated reporter gene expression in an anterior region of *D. melanogaster* wings despite the fact that adult *D. pseudoobscura* lack pigmentation in this area (Figure 2F, arrowhead). This expression pattern does not appear to be an artifact of the heterologous transgenic host because a similar pattern is seen in native *D. pseudoobscura* Yellow expression (see Figure 1 in (Gompel et al. 2005)). Interestingly, *D. tristis*, which is a member of the obscura group to which *D. pseudoobscura* also belongs, has a similar pattern of *yellow* expression in pupal wings controlled by an intronic enhancer and does display a corresponding spot of pigmentation on its adult wings (Prud'homme et al. 2006). This spot of wing pigmentation appears to be a derived trait in the obscura group (Prud'homme et al. 2006), thus the presence of this expression pattern in *D. pseudoobscura* suggests that the novel *yellow* enhancer activity in this wing spot preceded other changes, such as a coincident decrease in Ebony protein expression (Wittkopp et al. 2002a; Gompel et al. 2005), that are also required for wing spot formation.

Acknowledgements

We thank X.-W. Heng, E. Stewart, and A. Neidert for technical assistance; the Biological Resource Branch at the National Cancer Institute for recombineering reagents and S. Warming for recombineering advice; M. Calos for plasmids containing attB sequence; N. Gompel for advice on wing dissections and pictures of the six Drosophila species; J. Gruber, J. Coolon, A. Cooley, L. Arnold, D. Yuan, and B. Metzger for helpful comments on the manuscript; and T. Werner and S. B. Carroll for sharing unpublished data.

Figure 2-1

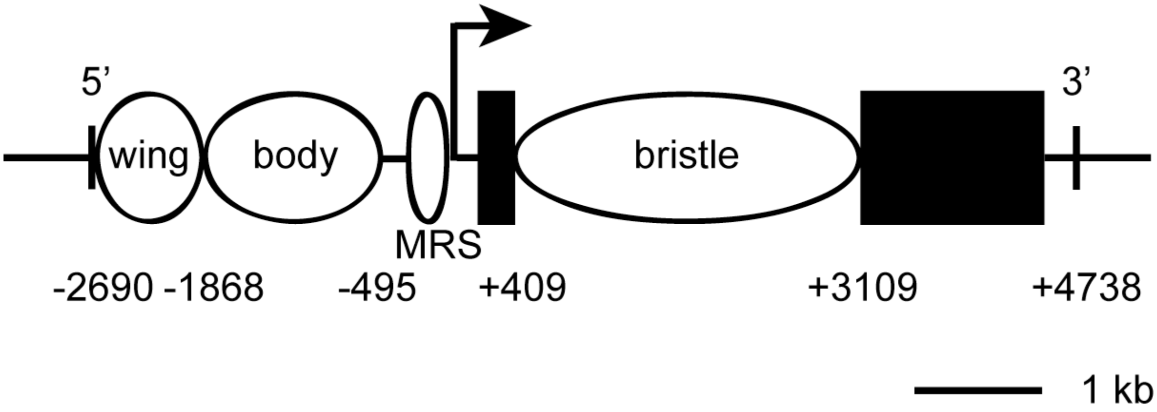


Figure 2-1. The *Drosophila melanogaster yellow* gene is regulated by multiple, tissue-specific enhancers.
The 5' intergenic region contains enhancers (open ovals) that drive expression in the wing and body of adult flies (Geyer and Corces 1987a; Martin et al. 1989b; Wittkopp et al. 2002d; Jeong et al. 2006b) as well as sequences known to influence male mating success (MRS, (Drapeau et al. 2006b)). It also contains sequences necessary for expression of *yellow* in the larval mouthparts, larval, denticle belts, microsatae, tarsal claws and sex combs (Geyer and Corces 1987a; Martin et al. 1989b). The intron contains the bristle enhancer as well as sequences necessary for *yellow* expression in larval mouthparts and larval denticle belts, tarsal claws, sex combs and aristae (Geyer and Corces 1987a; Martin et al. 1989b). Solid black boxes indicate the two exons of *yellow*, the arrow indicates the transcription start site, and vertical black lines indicate the 5' and 3' ends of sequence shown to fully rescue a *D. melanogaster yellow* mutant (Geyer and Corces 1987a).

Figure 2-2

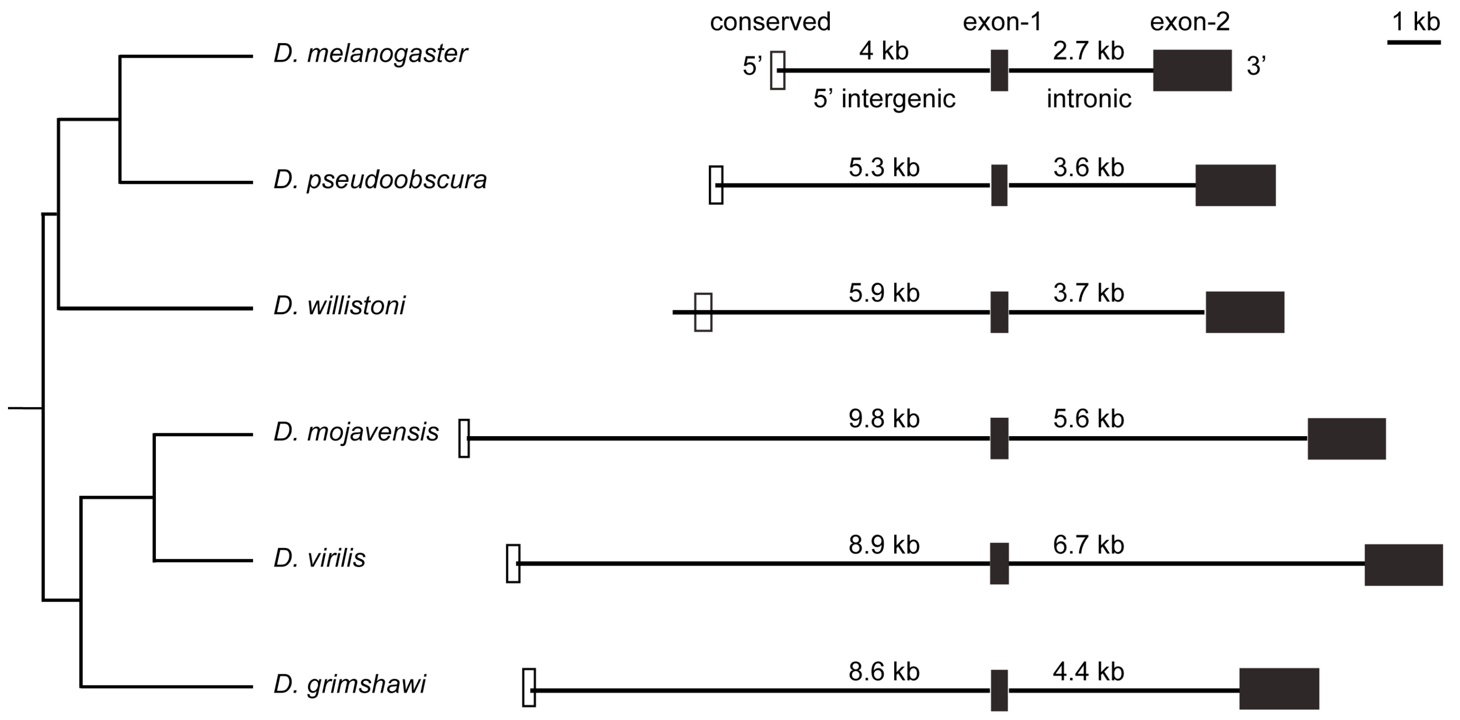

conserved
exon-1
exon-2
1 kb
5'
4 kb
2.7 kb
3'
5' intergenic
intronic
D. melanogaster
D. pseudoobscura
5.3 kb
3.6 kb
D. willistoni
5.9 kb
3.7 kb
D. mojavensis
9.8 kb
5.6 kb
D. virilis
8.9 kb
6.7 kb
D. grimshawi
8.6 kb
4.4 kb

Figure 2-2. DNA sequences tested for enhancer activity vary in length among species.
The size of each 5’ intergenic and intronic region tested, which ended and began, respectively at exon 1, is shown in kilobases (kb). Filled black boxes indicate exons, whereas open boxes indicate the region of conserved sequence shown in Figure 2-6. The black lines indicate the DNA included in each construct. Note that only *D. willistoni* includes the entire conserved 5’ block. Phylogenetic relationships among Drosophila species are indicated on the left.

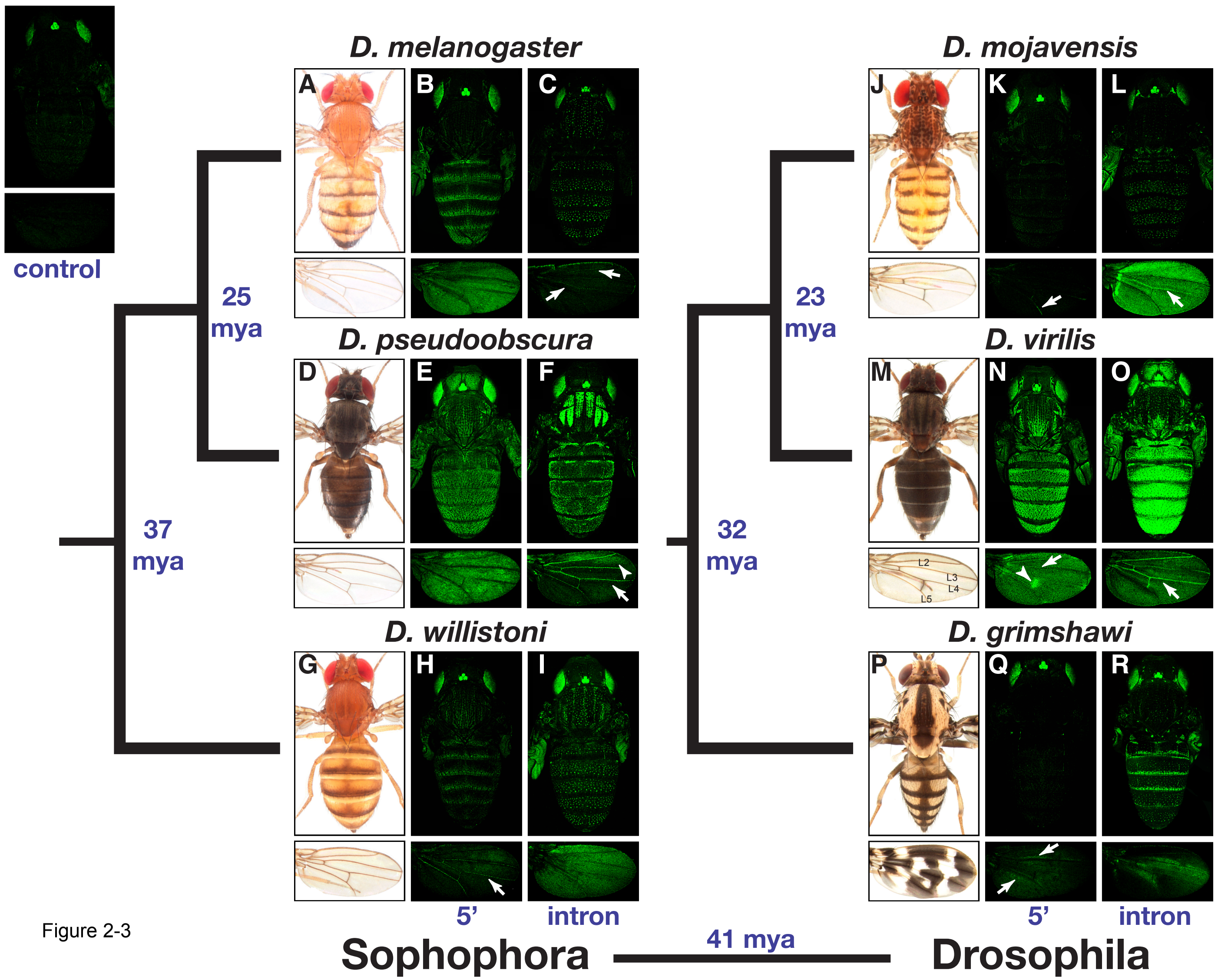

control
37
mya
25
mya
D. melanogaster
A
B
C
D. pseudoobscura
D
E
F
D. willistoni
G
H
I
5’
intron
Sophophora
41 mya
Drosophila
32
mya
23
mya
D. mojavensis
J
K
L
D. virilis
M
N
O
L2
L3
L4
L5
D. grimshawi
P
Q
R
5’
intron

Figure 2-3

Figure 2-3. Location and activity of the *yellow* body and wing enhancers is highly divergent among Drosophila species.
Expression (shown in green) of nuclear Green Fluorescent Protein (GFP) activated in transgenic *D. melanogaster* by the 5' intergenic (5') and intronic (intron) fragments of DNA shown in Figure 2 from the six species indicated is shown. For each species, the panel of six images includes pictures of the dorsal side of the body (top) and wings (bottom). From left to right, panels show an adult specimen of the species indicated (images provided by N. Gompel), a transgenic *D. melanogaste*r pupa carrying the corresponding 5' intergenic sequence-GFP reporter gene, and a transgenic *D. melanogaste*r pupa carrying the corresponding intronic sequence-GFP reporter gene. Expression patterns indicated with arrows and arrowheads are described in the main text. Divergence times (Russo et al. 1995; Spicer and Bell 2002) between lineages are shown in blue in millions of years ago (mya). Fluorescence observed in the body (top) and wing (bottom) of a *D. melanogaster* pupa carrying the GFP reporter gene without any putative *yellow* enhancer sequences cloned upstream is shown in the top left of the figure, and serves as a negative control. In each case, the GFP-expressing image shown is from female pupae, 70-80 hours old, and is representative of the at least 10 individual specimens examined of each genotype. Note that bright GFP expression in eyes and ocelli (located between eyes on each head) in all images, including the control, is activated by the transformation marker gene and not the y*ellow* 5' intergenic or intronic sequences.

Figure 2-4

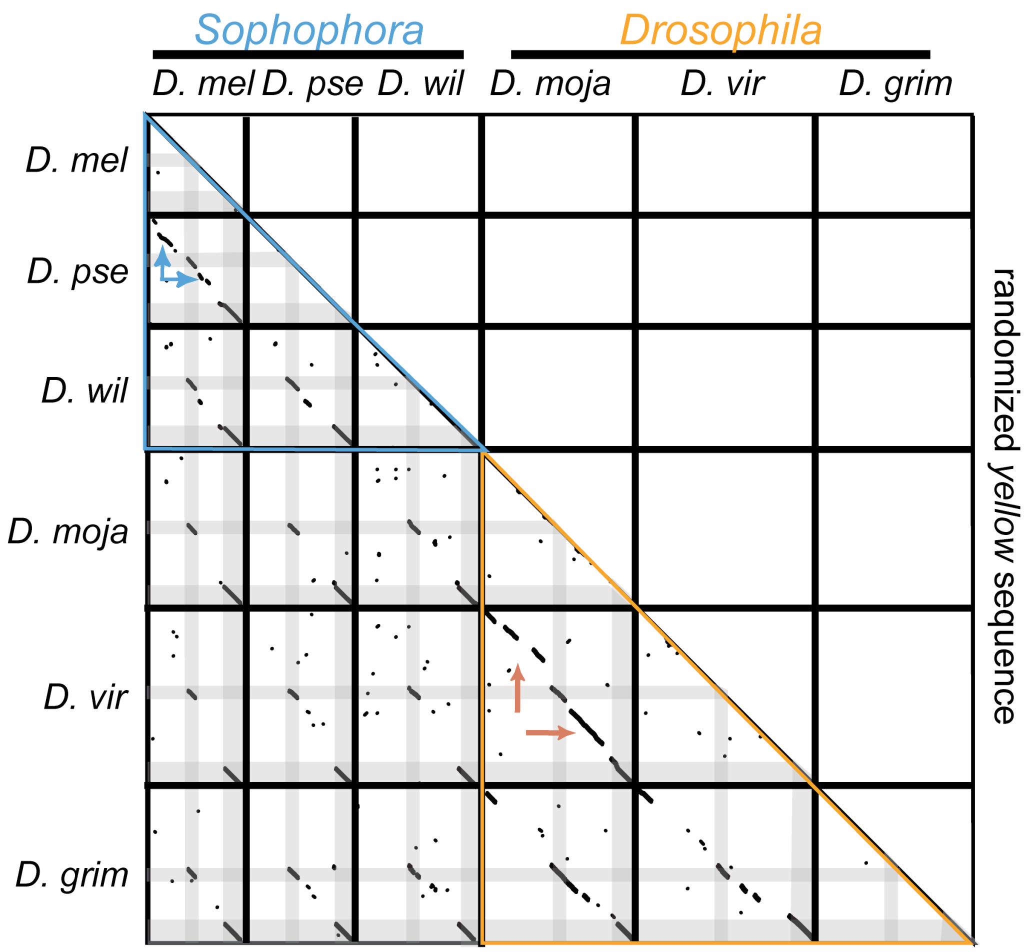

Sophophora
Drosophila
D. mel
D. pse
D. wil
D. moja
D. vir
D. grim
randomized *yellow* sequence

Figure 2-4. *yellow* sequences show no evidence of large duplications or transpositions.

Pairwise comparisons of *yellow* genes and their associated 5’ intergenic regions from each species to each other species (and to themselves) are shown in the lower left, and a comparison of each species’ sequence to a randomized version of these sequences is shown in the upper right. Sequence of each gene is from 5’ to 3’ from left to right and from top to bottom. Solid black lines separate one species’ sequence from the next, and regions corresponding to sequences from exon 1 and exon 2 are shaded grey in the lower left half. Comparisons among species within the subgenus Sophophora are outlined in blue, whereas comparisons among species within the subgenus Drosophila are outlined in orange. The remaining black pixels indicate blocks of sequence similarity identified using LASTZ, as described in the Materials and Methods. The red and blue arrows indicate regions of collinear sequence similarity discussed in the main text.

Figure 2-5

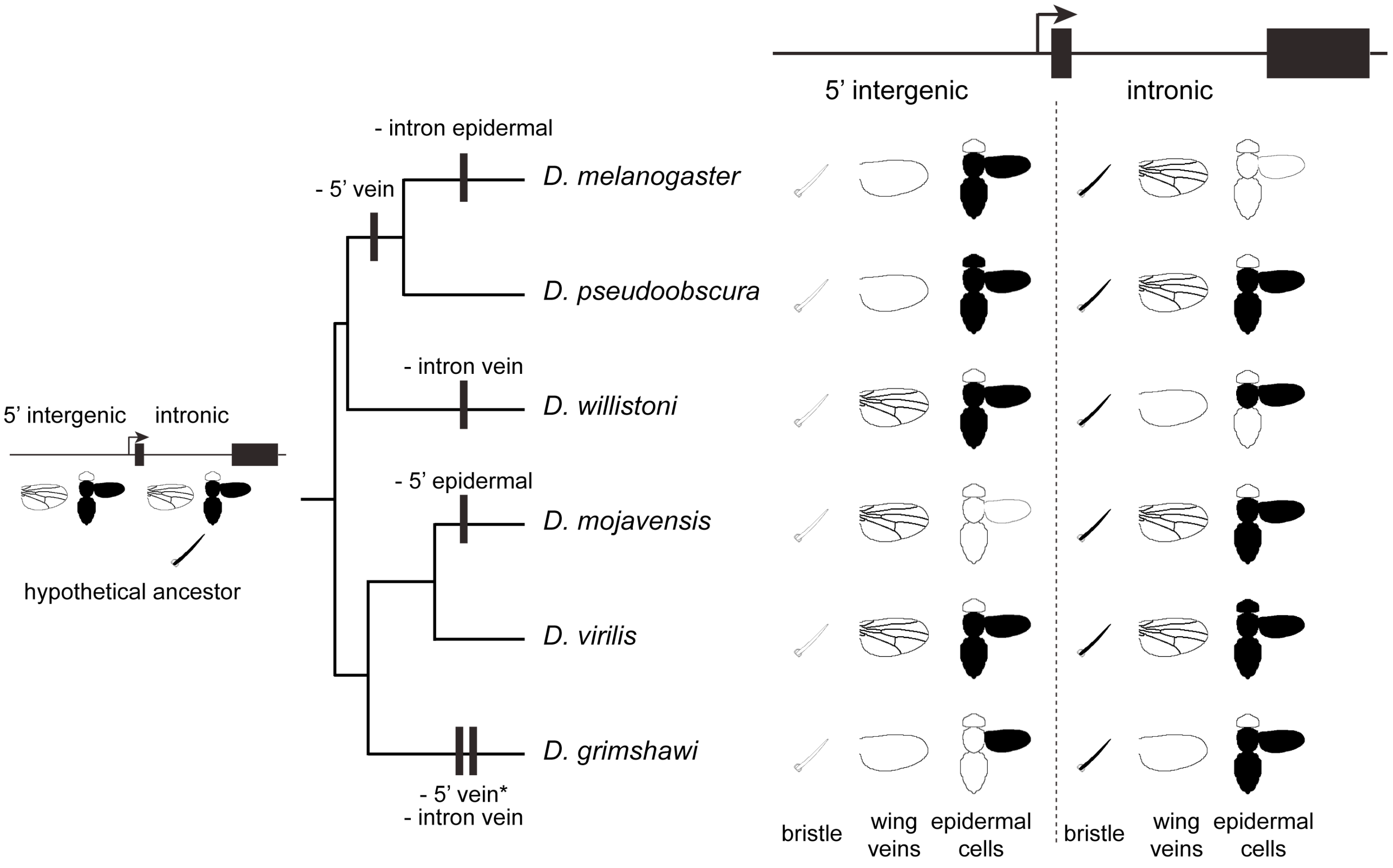
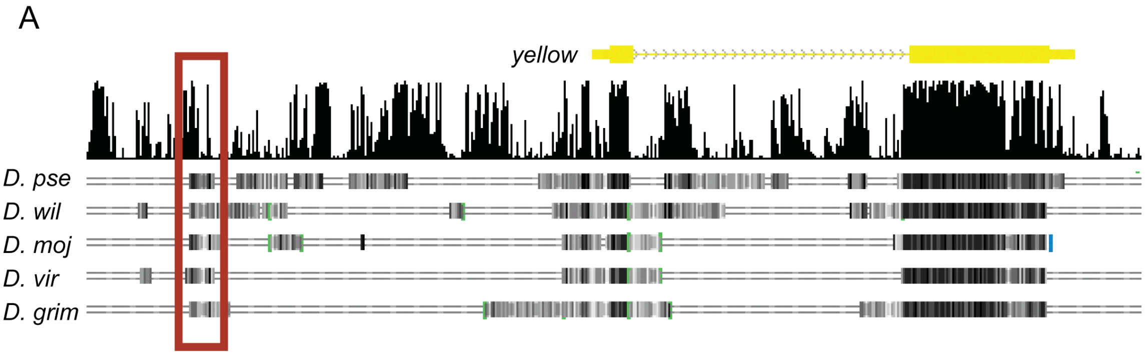

5’ intergenic
intronic
hypothetical ancestor
- intron epidermal
- 5’ vein
- intron vein
- 5’ epidermal
- 5’ vein*
- intron vein
D. melanogaster
D. pseudoobscura
D. willistoni
D. mojavensis
D. virilis
D. grimshawi
5’ intergenic
intronic
bristle
wing veins
epidermal cells
bristle
wing veins
epidermal cells

Figure 2-5. Dynamic *yellow cis*-regulatory architecture among Drosophila species.
The schematic summarizes enhancer activity of 5’ intergenic and intronic sequences from each of the six species shown. In the bristle and epidermal cell schematics - the latter of which shows a head, thorax, abdomen, and wing - regions shaded in black showed GFP expression. For the wing vein schematics, pictures of wings including visible veins indicate vein enhancer activity. A phylogenetic tree showing the relationship among species is shown to the left of the enhancer expression summary. A hypothetical *cis*-regulatory architecture of the common ancestor of these six species is shown with wing vein and epidermal cell enhancers in both the 5’ intergenic and intronic regions. Vertical black bars on the branches of the phylogenetic tree indicate losses of enhancer activity. The asterisk next to “- 5’ vein activity” is because unpublished data from T. Werner and S.B. Carroll indicates that an enhancer driving expression in the wing veins (as well as additional wing epidermal cells) is located upstream of the 5’ intergenic region we examined; no information about the expression of this reporter gene in epidermal cells of the head, thorax, or abdomen was available.

Figure 2-6



B

```
  D. melanogaster  acagtcggtttgtagttgc-gatttaaatagtttttaattaataactaattgcagcgaga-------cga
D. pseudoobscura  --------ttgacagttgc--atttaaatagtttttaattgaaaactaattgccacgagt-------gga
   D. willistoni  ---------------tttc-agtgcaattagttttcaattaaaaactatttaaatcgcaaccgattccga
   D. mojavensis  acaactcgttagcagctgctgatttaaatagtttttaattgccagcaaattg-agccgaa-------taa
      D. virilis  ccaactcgttagcagctggcgatttaaatagtttttaattggcagctaattg-agctgaa-------tag
     D. grimshawi  ----------------------tttaaatagtatttgtatggaaactaattg-agctgaa-------cga

  D. melanogaster  gtgcca-----------aa----acaagaca-g----tggtcgaca-------------tagctgtga-c
D. pseudoobscura  gtgcca-----------aa----gcgcgacg-acagctggacgaca-------------tggttgtga-c
   D. willistoni  ttgccagtcagactcccaa----gcaatccg-g----caaacgccacagaaactataactgaaaatga-a
   D. mojavensis  gtgcca-----------aa----actggacaag----tgcgtggca-----------gctggccgaca-c
      D. virilis  ctgcca-----------aa----actggacaag----tgcgtggca-----------actggccgaca-c
     D. grimshawi  gttcca-----------aacaggaaagaaaatg----tacgcgata-----------------catcagc

  D. melanogaster  gttttcaatgcag------caacatggtaacatggcaactat-tt------------------------a
D. pseudoobscura  gtttccagctctc------------------gctgcaacgctctt------------------------a
   D. willistoni  aactcaaaaacaa---aaccaaactagttaaatgcgaaatat-ttctctcttgggtttcatttccttcac
   D. mojavensis  gtcttgtcgtctc------t--------------------------------------------------
      D. virilis  gtcttgtcgtctc------tta---------agccaaacagt-gt----------------------ggg
     D. grimshawi  atcttatcgtctc------tcgt-tatttgtattcaagccac-tt----------------------aaa

  D. melanogaster  gcca-gtt------------aattggcaa--caa-gccgacagcgatatt----------------agca
D. pseudoobscura  gccacttt------------aattggcag--cag---cagcagcagccaa----------------aaca
   D. willistoni  gcca-tttggaaattcttggaattgcata--caa--ttggcagaaataat----------------aata
   D. mojavensis  ------------------------------------------------ca---------------cacaa
      D. virilis  gcac-ac---------------acaacaa--cagcaacaacaacaatcca---------------cacaa
     D. grimshawi  tgcc-acc------------agctgccaa--taccaccggcacccatccg--------------ttgctg

  D. melanogaster  gcagcaacaacagcttt--tcaccaggaaatcaata-----atagtttgtccaaaa--cata------at
D. pseudoobscura  gcggcaacaacaacttt--tcaccaggaaatcggca-----a---atgatccgaaagccaaacgagtcgc
   D. willistoni  ataataataataattatgagcataaataagtagata-----gtaaaggaaacaaat--aaaa------ct
   D. mojavensis  gagaca---gcaacttt--tcaccaggaaatcgaga-----a----------aacattcaaa------aa
      D. virilis  ccaacaatgacaacttt--tcaccaggaaaccaagaaaaaca----------aaaatacaaa------at
     D. grimshawi  gcagtactgacaacttt--tcaccaggaagccgaca-----a----------gaaaacaaaa------aa

  D. melanogaster  aataaaataaagccg--
D. pseudoobscura  aataaccgaaagctg--
   D. willistoni  gacacacacacacag--
   D. mojavensis  caaaaaca---------
      D. virilis  caaaaaaa---------
     D. grimshawi  aagagact-----tgca
```

Figure 2-6. Conserved region of non-coding sequence defines an orthologous endpoint for 5' intergenic regions.

(A) A schematic of the *yellow* gene is shown in yellow in which arrowheads point toward 3' end of the gene, thicker yellow boxes indicate the protein coding sequences with the two exons, and narrower yellow boxes indicate the 5‘ and 3‘ UTRs. Below this image is a histogram representing the extent of sequence conservation among 12 Drosophila species, mosquito, honeybee, and beetle, as determined using a Multiz alignment (Blanchette et al. 2004) and phastCons Scores (Siepel et al. 2005) and reported on the *D. melanogaster* UCSC Genome Browser (Rhead et al., 2010), (http://genome.ucsc.edu/). The region shown is located on the X chromosome and extends from position 245,638 to 258,882 in the April 2006 (BDGP R5/dm3) assembly. Taller bars indicate greater sequence conservation. Below this histogram is a density plot indicating the amount of sequence conservation between each species and *D. melanogaster*; darker bars indicate higher degrees of conservation, as scored by phastCons (Siepel et al. 2005). Vertical green and blue lines in these density plots indicate a lack of collinearity with *D. melanogaster*. The red box indicates the conserved region used to determine an orthologous 5' end to the intergenic fragments tested. (B) An alignment of sequences from the species examined in this study is shown for the boxed conserved region, which extends from positions 246,638 to 246,882 in the *D. melanogaster* genome (April 2006 (BDGP R5/dm3) assembly). Dashes indicate insertions or deletions among the twelve Drosophila species and honeybee sequence.

Figure 2-7

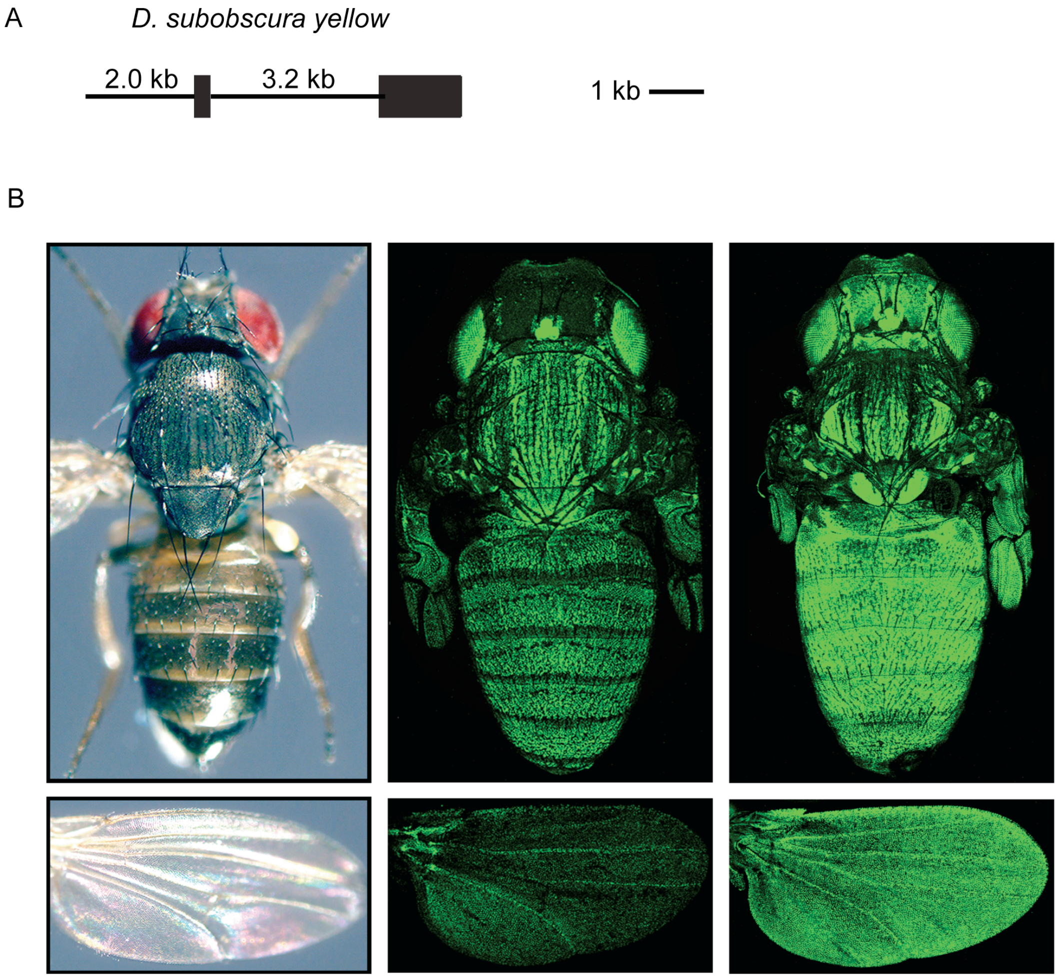


Figure 2-7. *D. subobscura* 5' intergenic and intronic *yellow* sequences both contain epidermal cell enhancers.

(A) A schematic of the *D. subobscura yellow* gene is shown with the amount of 5' integenic (2.0 kb) and intronic (3.2 kb) DNA included in the reporter genes indicated. (B) Images of dorsal bodies (top row) and wings (bottom row) from an adult wild-type *D. subobscura* (left) and *D. melanogaster* transformant pupae carrying a GFP reporter gene controlled by sequences from the 5' intergenic (middle) or intronic (right) region of *D. subobscura yellow* shown in (A). Like *D. pseudoobscura*, its closest relative among the species surveyed, expression in epidermal cells of the wing, abdomen, and thorax is driven by both the 5' intergenic and intronic regions. Overall, the pattern of expression is similar between the two species, although some differences are apparent. For example, expression in the head cuticle is driven by intronic sequences from *D. subobscura*, but 5' intergenic sequence from *D. pseudoobscura*; the 5' intergenic region of *D. subobscura* drives expression in the wing veins whereas the *D. pseudoobscura* 5' intergenic region does not; and the *D. subobscura* intron lacks the elevated spot of expression in the anterior part of the wing seen in *D. pseudoobscura*.

Figure 2-8

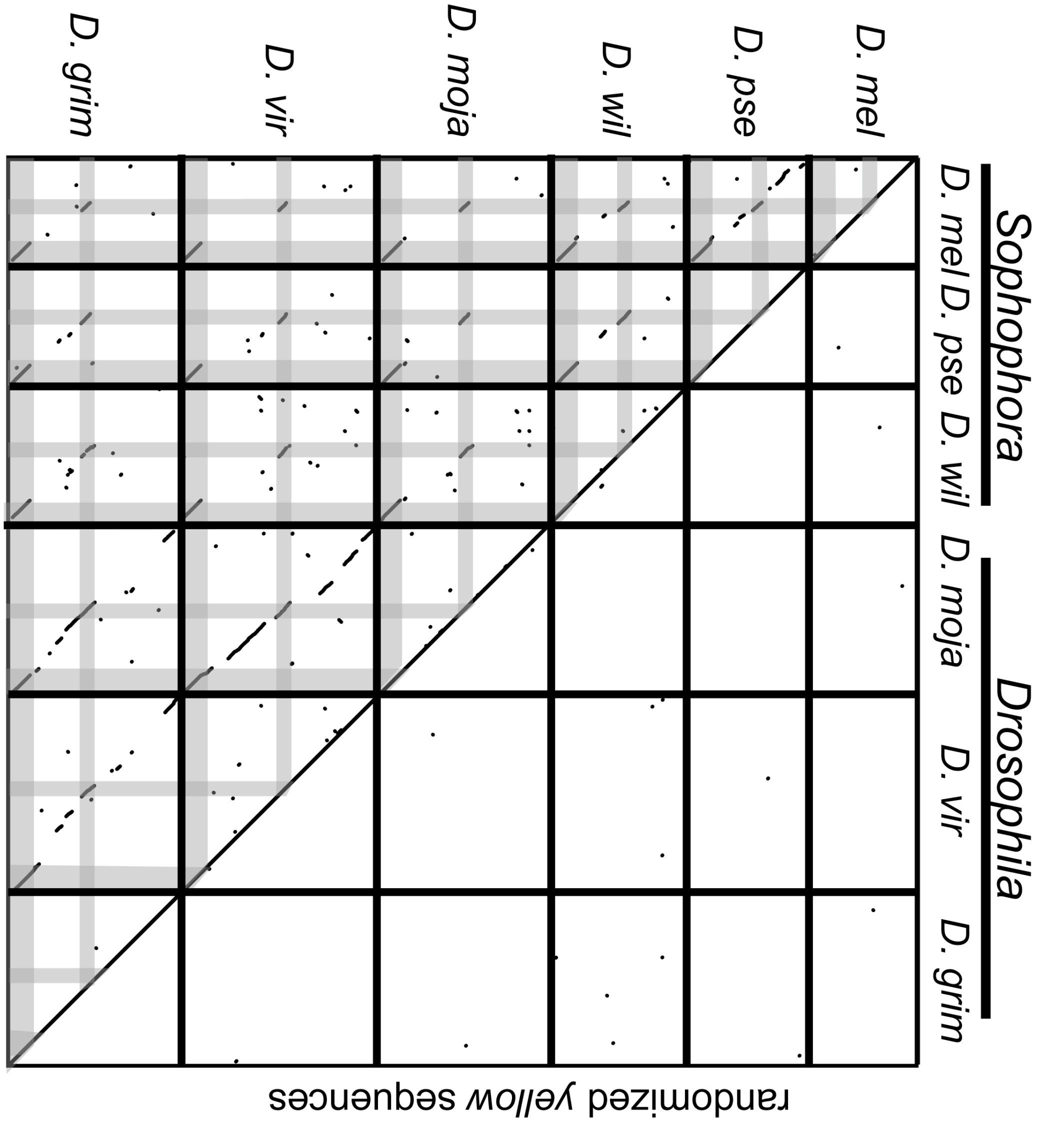


Figure 2-8. Alternative sequence alignment parameters also show primarily collinear sequence similarity.
Relaxing alignment parameters identified more regions of sequence similarity between species, but still showed no evidence of large duplications or transpositions. Figure format is as described in the legend to Figure 4 in the main text, and analysis conditions are as described in the Materials and Methods.

Table 2-1: BAC clones containing *yellow* and flanking genes

Species	BAC clone number[1]	Sequences included in the BAC clone				
		CG3777	consv[2]	*yellow*	*achete*	*CG4165*
D. pseudoobscura	CH222-46B19	-	NA[3]	+	-	NA
	CH222-11G23	-	NA	+	-	NA
	CH222-11H21	-	NA	+	-	NA
D. willistoni	DW18M19	-	NA	-	-	NA
	DW14D14	-	NA	+	+	NA
	DW10L5	+	+	+	+	NA
D. mojavensis	DM32A21	+	+	+	NA	+
	DM14M10	+	+	+	NA	+
	DM4J24	+	+	+	NA	+
	DM1M5	+	+	+	NA	-
	DM1C16	+	NA	-	NA	-
	DM9D16	+	+	+	NA	+
	DM2K5	+	NA	-	NA	-
	DM31I1	+	NA	-	NA	-
	DM9O22	+	+	+	NA	-
D. virilis	DV98O14	-	+	+	+	NA
	DV106J1	-	+	+	+	NA
	DV1A7	-	+	+	+	NA
	DV50B1	-	+	+	+	NA
	DV2K10	-	+	+	+	NA
	DV30O8	-	+	+	+	NA
	DV88F9	-	+	+	+	NA
	DV22J8	-	+	+	+	NA
	DV52F20	-	-	+	+	NA
	DV124G7	-	-	+	+	NA
	DV20J7	-	+	+	+	NA
	DV136F3	-	+	+	+	NA
D. grimshawi	DG8L1	+	+	+	+	NA
	DG18J1	+	+	+	+	NA
	DG27E24	+	+	+	+	NA
	DG41C3	+	+	+	+	NA
	DG3B4	+	+	+	+	NA
	DG3M22	+	+	+	+	NA
	DG8J4	-	NA	+	+	NA
	DG38N3	-	NA	+	+	NA
	DG4G9	-	NA	-	-	NA
	DG23K7	+	+	+	+	NA

[1] BAC clones used for reporter gene construction are shown in bold

[2] conserved region of sequence 5' of *yellow* shown in Supplementary Figure 1

[3] NA=Not Available (ie., untested)

References

Arakane, Y., N. T. Dittmer, Y. Tomoyasu, K. J. Kramer, S. Muthukrishnan, R. W. Beeman, and M. R. Kanost. Identification, mRNA expression and functional analysis of several yellow family genes in Tribolium castaneum. Insect Biochem Mol Biol 40:259-266.

Bischof, J., R. K. Maeda, M. Hediger, F. Karch, and K. Basler. 2007. An optimized transgenesis system for Drosophila using germ-line-specific phiC31 integrases. Proc Natl Acad Sci U S A 104:3312-3317.

Cande, J., Y. Goltsev, and M. S. Levine. 2009. Conservation of enhancer location in divergent insects. Proc Natl Acad Sci U S A 106:14414-14419.

Cretekos, C. J., Y. Wang, E. D. Green, J. F. Martin, J. J. t. Rasweiler, and R. R. Behringer. 2008. Regulatory divergence modifies limb length between mammals. Genes Dev 22:141-151.

Drapeau, M. D., S. A. Cyran, M. M. Viering, P. K. Geyer, and A. D. Long. 2006. A cis-regulatory sequence within the yellow locus of Drosophila melanogaster required for normal male mating success. Genetics 172:1009-1030.

Frankel, N., G. K. Davis, D. Vargas, S. Wang, F. Payre, and D. L. Stern. 2010. Phenotypic robustness conferred by apparently redundant transcriptional enhancers. Nature:May 30. [Epub ahead of print].

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

Geyer, P. K., and V. G. Corces. 1987. Separate regulatory elements are responsible for the complex pattern of tissue-specific and developmental transcription of the yellow locus in Drosophila melanogaster. Genes Devel. 1:996-1004.

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

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

Hare, E. E., B. K. Peterson, V. N. Iyer, R. Meier, and M. B. Eisen. 2008. Sepsid even-skipped enhancers are functionally conserved in Drosophila despite lack of sequence conservation. PLoS Genet 4:e1000106.

Helms, A. W., A. L. Abney, N. Ben-Arie, H. Y. Zoghbi, and J. E. Johnson. 2000. Autoregulation and multiple enhancers control Math1 expression in the developing nervous system. Development 127:1185-1196.

Hong, J. W., D. A. Hendrix, and M. S. Levine. 2008. Shadow enhancers as a source of evolutionary novelty. Science 321:1314.

Horn, C., and E. A. Wimmer. 2000. A versatile vector set for animal transgenesis. Dev Genes Evol 210:630-637.

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

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

Jeong, Y., K. El-Jaick, E. Roessler, M. Muenke, and D. J. Epstein. 2006b. A functional screen for sonic hedgehog regulatory elements across a 1 Mb interval identifies long-range ventral forebrain enhancers. Development 133:761-772.

Langeland, J. A., and S. Carroll. 1993. Conservation of regulatory elements controlling hairy pair-rule stripe formation. Development 117:585-596.

Ludwig, M. Z., N. H. Patel, and M. Kreitman. 1998. Functional analysis of eve stripe 2 enhancer evolution in *Drosophila*: rules governing conservation and change. Development 125:949-958.

Lukowitz, W., C. Schroder, G. Glaser, M. Hulskamp, and D. Tautz. 1994. Regulatory and coding regions of the segmentation gene hunchback are functionally conserved between Drosophila virilis and Drosophila melanogaster. Mech Dev 45:105-115.

Markstein, M., C. Pitsouli, C. Villalta, S. E. Celniker, and N. Perrimon. 2008. Exploiting position effects and the gypsy retrovirus insulator to engineer precisely expressed transgenes. Nat Genet 40:476-483.

Martin, M., Y. B. Meng, and W. Chia. 1989. Regulatory elements involved in the tissue-specific expression of the yellow gene of Drosophila. Mol. Gen. Genet. 218:118-126.

Pan, D., S. A. Valentine, and A. J. Courey. 1994. The bipartite D. melanogaster twist promoter is reorganized in D. virilis. Mech Dev 46:41-53.

Pappu, K. S., E. J. Ostrin, B. W. Middlebrooks, B. T. Sili, R. Chen, M. R. Atkins, R. Gibbs, and G. Mardon. 2005. Dual regulation and redundant function of two eye-specific enhancers of the Drosophila retinal determination gene dachshund. Development 132:2895-2905.

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

Rhead, B., D. Karolchik, R. M. Kuhn, A. S. Hinrichs, A. S. Zweig, P. A. Fujita, M. Diekhans, K. E. Smith, K. R. Rosenbloom, B. J. Raney, A. Pohl, M. Pheasant, L. R. Meyer, K. Learned, F. Hsu, J. Hillman-Jackson, R. A. Harte, B. Giardine, T. R. Dreszer, H. Clawson, G. P. Barber, D. Haussler, and W. J. Kent. The UCSC Genome Browser database: update 2010. Nucleic Acids Res 38:D613-619.

Romano, L. A., and G. A. Wray. 2003. Conservation of Endo16 expression in sea urchins despite evolutionary divergence in both cis and trans-acting components of transcriptional regulation. Development 130:4187-4199.

Russo, C. A., N. Takezaki, and M. Nei. 1995. Molecular phylogeny and divergence times of drosophilid species. Mol Biol Evol 12:391-404.

Sambrook, J., and D. Russell. 2001. Molecular Cloning: a Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor.

Schwartz, S., W. J. Kent, A. Smit, Z. Zhang, R. Baertsch, R. C. Hardison, D. Haussler, and W. Miller. 2003. Human-mouse alignments with BLASTZ. Genome Res 13:103-107.

Spicer, G. S., and C. D. Bell. 2002. Molecular Phylogeny of the Drosophila virilis Species Group (Diptera: Drosophilidae) Inferred from Mitochondrial 12S and 16S Ribosomal RNA Genes. . Annuals of the Entomological Society of America 95:156-161.

Stothard, P. 2000. The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. Biotechniques 28:1102, 1104.

Team, R. D. C. 2005. A language and environment for statistical computing. . R Foundation for Statistical Computing, Vienna, Austria.

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

Uemura, O., Y. Okada, H. Ando, M. Guedj, S. Higashijima, T. Shimazaki, N. Chino, H. Okano, and H. Okamoto. 2005. Comparative functional genomics revealed conservation and diversification of three enhancers of the isl1 gene for motor and sensory neuron-specific expression. Dev Biol 278:587-606.

Werner, T., S. Koshikawa, T. M. Williams, and S. B. Carroll. 2010. Generation of a novel wing colour pattern by the Wingless morphogen. Nature 464:1143-1148.

Wittkopp, P. J. 2006. Evolution of cis-regulatory sequence and function in Diptera. Heredity 97:139-147.

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

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

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

References for Supporting Information

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

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

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

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

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

Werner, T., S. Koshikawa, T. M. Williams, and S. B. Carroll. 2010. Generation of a novel wing colour pattern by the Wingless morphogen. Nature 464:1143-1148.

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

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

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

Siepel, A., G. Bejerano, J. S. Pedersen, A. S. Hinrichs, M. Hou, K. Rosenbloom, H. Clawson, J. Spieth, L. W. Hillier, S. Richards, G. M. Weinstock, R. K. Wilson, R. A. Gibbs, W. J. Kent, W. Miller, and D. Haussler. 2005. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Res 15:1034-1050.

Rhead, B., D. Karolchik, R. M. Kuhn, A. S. Hinrichs, A. S. Zweig, P. A. Fujita, M. Diekhans, K. E. Smith, K. R. Rosenbloom, B. J. Raney, A. Pohl, M. Pheasant, L. R. Meyer, K. Learned, F. Hsu, J. Hillman-Jackson, R. A. Harte, B. Giardine, T. R. Dreszer, H. Clawson, G. P. Barber, D. Haussler, and W. J. Kent. The UCSC Genome Browser database: update 2010. Nucleic Acids Res 38:D613-619.

Blanchette, M., W. J. Kent, C. Riemer, L. Elnitski, A. F. Smit, K. M. Roskin, R. Baertsch, K. Rosenbloom, H. Clawson, E. D. Green, D. Haussler, and W. Miller. 2004. Aligning multiple genomic sequences with the threaded blockset aligner. Genome Res 14:708-715.

Chapter 3

Dissecting *cis* and *trans* regulatory factors affecting *yellow* regulation in three Drosophila species

Abstract

Using the *yellow* gene and the changes in its enhancers as a model to understand how *cis*-regulatory elements change over evolutionary time, I previously found that the position and activity of *yellow* epidermal-cell enhancer has diverged rapidly among species. In order to understand the changes underlying this rapid evolution I conducted detailed comparisons between enhancer activities and the corresponding binding factors. Therefore, I subdivided *yellow* 5' intergenic and intronic regions from three Sophophora group species, *Drosophila melanogaster*, *D. pseudoobscura* and *D. willistoni* into 1-kilobasepair pieces. I then tested these sub-elements for enhancer activity using the Green Fluorescent Protein gene as the reporter and *D. melanogaster* as the common transgenic host. I found that some sub-elements drove expression that partially or fully recapitulated the expression pattern driven by the full region, whereas some failed to drive expression above background levels. Intriguingly, some sub-elements (~half of the total) drove epidermal-cell expression in spatial patterns that were not part of the pattern driven by the full region. This might suggest that the cryptic expression patterns observed when *yellow* enhancer

sub-elements were tested in isolation are repressed by the surrounding sequences when in the native genomic position. The existence of such cryptic epidermal-cell enhancer activities may have facilitated the rapid divergence of the activity and position of *yellow* epidermal-cell enhancers, as the 5' intergenic and intronic regions of *yellow* were already primed for epidermal-cell enhancer activity (i.e., had a collection of appropriate transcription factors for epidermal-cell enhancer activity). Using a yeast-one-hybrid assay, I identified a set of 204 transcription factors that appear to be binding to yellow enhancer sub-elements previously tested for enhancer activity. The transcription factors identified suggest that *yellow* is a target of the ecdysone pathway and is also regulated by genes that are important for bristle development, for expression throughout the X chromosome (where *yellow* is located), and even by a well-known tumor suppressor gene conserved between humans and flies. RNA interference experiments so far showed four of the identified transcription factors affecting pigmentation, and two of these are novel pigmentation genes. Comparing similarities between transcription factor binding profiles and enhancer activities revealed that *yellow* enhancer sub-elements that drive male-specific abdominal expression tend to be bound by similar sets of transcription factors. Further tests are necessary to make the list of transcription factors binding to *yellow* enhancers more comprehensive, to validate their direct binding, and to elucidate their functional effects on enhancer activity. Only then, using evolution of *yellow* enhancers, one can understand how changes in enhancer sequence, transcription factor binding and activity affect each other throughout evolution.

Introduction

Enhancers harbor a certain combination of transcription factor binding sites, with specific spacing and orientation, as the building blocks of the regulatory information encoded in them. This regulatory information controls the transcription of their corresponding gene in a time, place and quantity specific manner. In addition to this primary structure, groups of transcription factors within an enhancer act together to create modules that have independent enhancer activity (Yuh and Davidson 1996; Yuh et al. 1996; McGregor et al. 2007; Frankel et al. 2010a). Each module typically drives a complementary expression pattern to the rest of the modules, such that all modules together make up the expression pattern of the full enhancer, although it is possible for different enhancer modules to have partially (McGregor et al. 2007) or highly overlapping activities (e.g., shadow enhancers) (Frankel et al. 2010a).

Over time mutations can alter transcription factor binding sites which may or may not change the activities of modules as well as complete enhancers. Relatively few changes in enhancer sequence can lead to changes in activity and especially between closely related species one can identify the responsible nucleotide changes by comparing the sequences of two orthologous enhancers. Frankel and colleagues (Frankel et al. 2011) were able show that only 14 nucleotide changes within a 1-kb enhancer module of the *shaven baby* (*svb*) gene can explain majority of the expression difference in the dorsolateral larval epidermis cells and the resulting phenotypic difference observed in trichome number between *D. melanogaster* and *D. sechelia*. Similarly, Williams et al. (Williams et al. 2008) showed that the male-specific abdominal expression of the *bric-a-brac* (*bab*) gene observed in *D. melanogaster,* but not in *D. willistoni,* is mostly due to changes in the number, polarity and composition of few Abdominal-B and Doublesex binding sites within one enhancer module. It is important to note, however, that since typically multiple enhancer modules together control the complete expression pattern of a gene, it is not atypical to find combination of changes in multiple enhancer modules responsible for changes in gene expression (McGregor et al. 2007; Frankel et al. 2010a).

The relationship between enhancer sequence and activity, however, is more complex than the case mentioned above, such that enhancers showing virtually identical activities in the same *trans* environment can have highly dissimilar, even unalignable, sequences. For instance, the four enhancers of the *even-skipped* gene (stripe 3+7, stripe 2, stripe 4+6 and muscle-heart) from *D. melanogaster* versus a distantly related fly species, *Sepsis cynipsea,* drive virtually identical expression patterns in the same *trans* environment, but there are few if any similarities in their sequences(Hare et al. 2008c). This is partly due to the degeneracy of transcription factor binding sites and partly to gradual compensatory gain and loss of transcription factor binding sites, where the sequence of an enhancer changes but its activity stays the same. This type of enhancer evolution, unfortunately, makes it harder to separate the functional sequence changes from neutral ones and, hence, to shed light onto the currently unknown code underlying enhancers.

Another (complex) feature of enhancer evolution, besides the changes in the sequence and function of existing enhancers, is the evolution of novel enhancers. This can happen several ways. A gene can gain a novel expression pattern through the insertion of a transposable element with enhancer activity, promoter-switching, *de novo* formation of an enhancer, or co-option of an existing enhancer (to drive a novel expression pattern, e.g., in a new domain). Previously, spread of *cis*-regulatory information in the genome via transposable elements has been proposed to be the prevalent mechanism for evolution of new gene expression patterns (Britten and Davidson 1971). However, examples of all four aforementioned trajectories of gaining novel expression patterns have now been idetified and the current number of studies do not allow making an assessment about the prevalent mechanism for evolution of novel enhancers. It is notable, however, that in the recent years examples of *de novo* enhancer formation and co-option of existing enhancers to create novel expression patterns have shed more light onto the possible molecular mechanisms underlying these two trajectories and showed that under appropriate selective conditions few nucleotide changes can result in novel gene expression patterns (e.g., expression in a different tissue type).

Eichenlaub and Ettwiller (Eichenlaub and Ettwiller 2011) found that following whole genome duplication, ancestrally coding sequences lost their coding ability and acquired enhancer activities through few mutational changes in teleosts. The orthologous regions of these *de novo* enhancers in mammals preserved coding ability, but did not appear to have enhancer activity in teleost or mammals, confirming the novelty of these *cis*-regulatory elements. In general, *de novo* evolution of enhancer activities is hard to detect and confirm, which may be one reason why there are so few examples of it. Nevertheless, it is important to consider it as a viable possibility when looking at how gene expression changes over time.

Co-option of existing enhancers to create novel expression patterns has also been shown to occur through a handful of mutations. One of the best and most recent examples of enhancer co-option is the gain of optic lobe expression by the *D. santomea Neprilysin-1* (*Nep-1*) gene as compared to its sister species *D. yakuba* (Rebeiz et al. 2011a). In this case the novel optic lobe enhancer gained activity by co-opting the preexisting enhancers, central nervous system (CNS) and retinal field, in its vicinity, with which it shared half of its sequence. Only four nucleotide changes were sufficient to explain the majority of the gain of enhancer activity in the optic lobe.

All of the above case studies showing enhancers with highly dissimilar sequences (but highly similar activities), orthologous enhancers harboring few nucleotide changes (but significant differences in activity), and evolution of enhancers with novel activity, document the dynamism and complexity of how enhancer sequence and function change over evolutionary time. With these features of enhancers in mind, in order to get a comprehensive understanding of their evolution, one needs to combine sequence comparisons with functional tests to identify the enhancer activities lying in candidate genomic regions and to identify the identity, composition and affinity of the corresponding transcription factor binding sites.

Previously, I studied yet another feature of enhancers, their genomic position relative to the coding sequence, in the Drosophila *yellow* gene. Genomic organization of enhancers are typically assumed to be conserved between species partly due to bias against publishing negative results like the failure to find enhancer conservation. This assumption is not always true: in the literature there are a few examples of enhancer position change(Pan et al. 1994b; Sanges et al. 2006); however, the assumption about conservation of enhancer position had not been explicitly and systematically tested before.

Using reporter transgenes, I identified the regions harboring epidermal-cell, wing-vein and bristle enhancers, of the *yellow* gene from six Drosophila species (*D. melanogaster*, *D. pseudoobscura*, *D. willistoni*, *D. mojavensis*, *D. virilis*, *D. grimshawi*). These enhancers primarily drive expression during the late pupal stage and are important for the pigmentation of the tissue types where they drive expression.

One of the most intriguing findings of this study was the high variability observed in the positions of the epidermal-cell and wing-vein enhancers with respect to the coding sequence. The locations of these two enhancers seemed to have changed between the intron and the 5’ intergenic region of *yellow* multiple times in the Drosophila evolutionary history. In some species both 5’ intergenic and intronic regions of *yellow* seemed to have epidermal-cell and wing-vein enhancer activity, whereas in other species these activities were located in one of the two regions only. In contrast, bristle enhancer was located in the intron of *yellow* in all six species investigated. Subsequent sequence comparisons showed no sign of a transposition or duplication event in or around *yellow* between closely related species that may have lead to the enhancer position changes observed. This suggested that these *yellow* enhancers must have been gained and lost between 5' intergenic and intronic regions via gradual compensatory gain and loss transcription factor binding sites multiple times int he evolutionary history.

The other important result was the divergence of the spatial pattern and strength of expression driven by *yellow* epidermal-cell enhancers which was correlated with the diverse cuticle pigmentation observed in species the enhancers were isolated from.

Hence, overall, I observed an abundance of diversity in the sequence, position and activity of two *yellow* enhancers among Drosophila species.

In order to elucidate how *yellow cis*-regulatory architecture changed rapidly over evolutionary time one needs to be able to do more detailed comparisons between sequence, activity and transcription factor binding profiles of *yellow* enhancers among Drosophila species. The size of the previously tested 5' intergenic and intronic regions of *yellow* are significantly bigger than a typical enhancer. To identify the specific changes causing altered enhancer activity and/or position, functional and sequence comparisons among smaller DNA fragments are necessary. Hence, I dissected the 5' intergenic and intronic regions of *yellow* from the Sophopora group species: *D. melanogaster*, *D. pseudoobscura* and *D. willistoni,* into sub-elements, tested them for enhancer activity as well as for their binding factors, and then compared these findings among sub-elements to understand overall how the *cis*-regulatory architecture of *yellow* has evolved.

As a result, I discovered that at the late pupal stage most sub-elements from the 5' intergenic regions of *yellow* showed epidermal cell enhancer activity, indicating that this activity is spread widely throughout the whole region. I found that some enhancer sub-elements drove expression patterns that almost recapitulate the expression driven by the full region they are isolated from. Some enhancer sub-elements, however, drove a fraction of the expression pattern coming from the full region, some sub-elements failed to drive expression above background levels and some drove expression patterns that were not part of the pattern driven by the full region they were isolated from suggesting they harbored cryptic enhancer activities. Finding cryptic activities was intriguing as they may have facilitated the observed rapid change in the activity and position of the *yellow* epidermal-cell enhancer.

I also identified a large set of candidate transcription factors binding to *yellow* 5' intergenic and intronic regions from the three Sophophora group fly species. I found that similar number of transcription factors interacted with *D. pseudoobscura* and *D. willistoni yellow* enhancers as it did with *D. melanogaster* ones suggesting that the transcription factor binding sites are mostly conserved between these three species. Only ~10-20% of the transcription factors were shared binders between regions from different species even though they harbored similar enhancer activities (e.g. epidermal-cell enhancer activity), which may be partly due to the high false negative rate of Y1H. I also found that despite the transcription factors that Y1H potentially missed, a set of enhancer sub-elements showing sexually dimorphic expression pattern in the abdomen, had high similarity in their transcription factor binding profiles. More detailed analysis of activities and transcription factor binding profiles of *yellow* enhancers can elucidate whether enhancers with similar expression patterns are bound by similar sets of transcription factors.

Overall, these results shed more light onto how *yellow cis*-regulatory architecture rapidly changed over evolutionary time and also lead the way to further investigations on how changes in the transcription factor binding profiles of *yellow* enhancers affect enhancer activity and position.

Results

yellow enhancer sub-elements were tested for activity

In order to understand how enhancer activities are distributed within the 5' intergenic and intronic regions of *yellow* from *D. melanogaster*, *D. pseudoobscura* and *D. willistoni,* I subdivided these regions into into approximately 1000 bp sub-elements (Figure 3-1). From 5'-to-3', each sub-element overlaps with the flanking ones by approximately 100 bp (Figure 3-1). I obtained 29 enhancer sub-elements to test for enhancer activity (using the reporter gene assay described in Chapter 2). In this test, nuclear enhanced green fluorescent protein (nEGFP), whose codons are optimized for expression in Drosophila, was used as the reporter and *D. melanogaster* was the common transgenic host. Enhancer activity was documented for transgenic lines that harbor 21 out of the 29 sub-elements of *yellow* enhancers from aforementioned three Drosophila species, as well as five transgenic lines that harbor full 5' intergenic or intronic regions of *yellow* and one negative control transgenic line that harbors the reporter gene with a basal promoter, but no putative enhancer, at the late pupal stage (70-80 hours after puparium formation (APF)) in individuals hemizygous for the transgene. In order to eliminate effects of chromosomal position on the expression of transgenes, attB-attP targeted insertion system was used to insert all transgenes in the same site on the chromosome arm 2L (25C6) (see Materials and Methods for full description of the genotype of the transgenic host). GFP expression in the eyes and ocelli was used to mark transformants, but was not recorded as part of the expression pattern driven by *yellow* enhancers. As a result, this experimental design allows comparison of reporter gene expression, and hence enhancer activity, among transgenic lines that harbor the full 5' intergenic or intronic regions of *yellow* from either of the three species as well as the transgenic lines that harbor the sub-elements of the 5' intergenic and intronic regions. Moreover, the expression differences observed between different transgenic lines are due to *cis*-regulatory differences between different enhancer regions tested since the *trans* environment, *D. melanogaster*, has been consistent among all transgenic lines.

How are enhancer activities distributed within the 5’ intergenic and intronic regions of *yellow*?

Enhancer activities can be distributed within the 5’ intergenic and intronic regions of *yellow* in several different patterns such that sub-elements within a region may harbor:

1- enhancer activity that recapitulates the expression patterns driven by the full region the sub-element is isolated from. (This can be referred to as the minimal enhancer.)

2- no activity above background levels.

3- enhancer activity that drives a fraction of the expression pattern driven by the full region the sub-element is isolated from.

4- cryptic enhancer activity driving additional expression not seen in the full element

Typically, in the gene regulation field, an enhancer region is thought to harbor sub-elements of the first three kinds mentioned above. This was true in the case of yellow 5’ intergenic and intronic regions as well, but additionally I discovered that all but one 5’ intergenic and intronic regions dissected also harbored at least one sub-element with cryptic activity, and overall, almost half of the sub-elements tested harbored some cryptic enhancer activity.

More specifically, as a result of analyzing activities of 21 enhancer sub-elements, I found that most, but not all, of the 1-kb-sub-elements tested so far showed enhancer activity in the epidermal cells in the abdomen, thorax or head or in bristle associated cells in the late pupal stage (Figure 3-2). Expression in the wing epidermal and vein cells was not documented in this analysis. I found that enhancer activities were mostly spread throughout the full DNA regions dissected, rather than being restricted to one or few sub-elements. Part or all of the spatial expression pattern driven by over one third (nine) of the sub-elements tested appeared to be a fraction of the enhancer activity driven by the full 5' intergenic or intronic region of *yellow.* On the other hand, about half (ten) of the sub-elements tested so far harbored cryptic activity such that the spatial expression pattern or the intensity of expression they drove in the epidermal cells was not a subset of the expression pattern driven by the intact 5' intergenic or intronic regions of *yellow* (Figure 3-2, asterisks). A little less than one third (six) of the sub-elements tested so far did not appear to show enhancer activity above background levels. Enhancer activities of individual sub-elements is described in more detailed below.

Enhancer activities lying in the sub-elements of 5' intergenic region of yellow *from* D. melanogaster

Among the five sub-elements from the 5' intergenic region of *yellow* from *D. melanogaster*, four showed enhancer activity in the epidermal cells in the late pupal stage, whereas one did not drive expression above background levels as compared to the negative control (Figure 3-2A). One out of the four sub-elements appeared to recapitulate the expression pattern driven by the full 5' intergenic region, albeit not completely (imperfect/incomplete minimal enhancer). One appeared to harbor fraction of the total expression pattern driven by the full region. Three sub-elements (including the one harboring a minimal enhancer) appeared to drive expression in spatial patterns that were not part of the pattern driven by the full region (cryptic enhancer activities). More detailed description of the expression patterns driven by D. melanogaster yellow 5' intergenic sub-elements are below. These sub-elements are shown in Figures 3-2A1, 2A2, 2A3, 2A4, 2A5 (from now on referred to as mel_A1, mel_A2, mel_A3, mel_A4, mel_A5 respectively).

Minimal enhancer mel_A2 drives expression in a spatial pattern that is highly reminiscent of the pattern driven by the full 5’ intergenic region of *yellow* from *D. melanogaster* where abdominal expression is observed mainly at the posterior end of each tergite as a horizontal stripe with a peak in the middle pointing towards the anterior (compare Figure 3-2A2 to 3-2A6). Moreover, both the full 5’ intergenic region of *D. melanogaster yellow* as well as mel_A2 drives sexually dimorphic expression in tergites A5 and A6, where in males, as compared to females, the posterior horizontal stripe with the middle peak is replaced by expression driven through out the width and length of both tergites. This type of male-specific abdominal *yellow* expression pattern correlates with and is one of the causes of the widely observed male-specific melanization patterns in A5 and A6. It is notable that as compared to the A5 and A6 expression the full element drives, mel_A2 enhancer activity does not completely fill these two tergites, which makes this enhancer sub-element an “incomplete minimal enhancer”. This may be because the breaking points of mel_A2 may have separated a functional element driving expression throughout A5 and A6. The rest of this activity may be located in one or more of the flanking sub-elements (mel_A1 and/or mel_A3) as well as in the other two (mel_A4 and mel_A5) that are not neighbors to mel_A2.

Cryptic enhancer activity Sub-element mel_A2 also appears to drive expression in the thorax, particularly in the scutellum, and head, which is not observed as a prominent part of the spatial expression pattern driven by the full 5’ intergenic region of *D. melanogaster yellow* (Figure 3-2A6). This can be regarded as “cryptic” enhancer activity and is observed also in sub-elements mel_A1 and mel_A3, where expression in the epidermal cells in the thorax and head is stronger as compared to the transgenic line harboring the full 5’ intergenic region of *D. melanogaster yellow* (in this transgenic line, expression in head epidermal cells appears to be absent). More noticeable cryptic expression patterns are driven by sub-elements mel_A1 and mel_A3, where reporter gene expression is observed in the epidermal cells throughout the length and width of each abdominal tergite, as opposed to the horizontal stripe at the posterior end of each of tergite as driven by the intact 5’ intergenic region of *D.melanogaster yellow*.

Fraction Sub-element mel_A4 appears to drive sexually dimorphic expression, such that in males expression driven in abdominal tergites A4, A5 and A6 appear stronger than the expression driven in the rest of the abdomen. Similarly, in females, mel_A4 appears to drive stronger expression in abdominal tergites A5 and A6 as compared to the rest of the abdomen. However, in this case the expression observed in the rest of the abdomen in females appear stronger than the expression observed in the corresponding tergites in males. The male specific expression pattern can be regarded as a fraction of the total expression pattern driven by the full 5’ intergenic region in male abdomens.

Enhancer activities lying in the sub-elements of 5’ intergenic region of yellow *from* D. pseudoobscura

Among the six *D. pseudoobscura yellow* 5’ intergenic sub-elements, enhancer activities of two are yet to be determined. Within the remaining four, one appears to drive expression in a spatial pattern highly reminiscent of the one driven by the full 5’ intergenic region (minimal enhancer). Two sub-elements appear to drive fraction of the total expression driven by the full region and one sub-element appears to drive cryptic expression in the epidermal cells encircling the head and with higher intensity, as compared to the expression driven by the full region, throughout the epidermal cells in thorax and abdomen.

Minimal enhancer pse_B1 drives expression in a spatial pattern that is highly similar to that of the spatial expression pattern driven by the full 5’ intergenic fragment (Figure 3-2B7), where epidermal cells along the length and width of each abdominal tergite show enhancer activity. Moreover, in both transgenic lines, abdominal tergites A5 and A6 show stronger expression in males than females, suggesting a sexually dimorphic spatial expression pattern. Both pse_B1 sub-element and the full *D. pseuodoobscura yellow* 5' intergenic region also drive expression in the epidermal cells on top of the head in a circle. The similarity between the thoracic expression pattern driven by these two regions is harder to asses due to virtually absent or weak activity observed in the two lines, respectively.

Fraction It is notable that the expression patterns driven by the sub-elements pse_B2 and pse_B3 can together recapitulate the expression pattern driven by the full 5' intergenic region of *D. pseudoobscura yellow*. Hence on their own they drive fraction of the expression driven by the full region, but together they can be regarded as a minimal enhancer. In this case, pse_B2 virtually lacks enhancer activity in the epidermal cells in the abdominal tergites A1, A2, A3, A4, but drives expression in A5 and A6, whereas pse_B3 virtually lacks enhancer activity in the epidermal cells in the abdominal tergites A5 and A6, but drives expression throughout the length and width of A1, A2, A3 and A4 as well as on the top of the head, complementing the expression pattern driven by pse_B2 to encompass the whole abdomen. Additionally, pse_B2 appears to drive sexually dimorphic expression throughout tergites A5 and A6 only in males; in females expression in these two tergites is restricted to their posterior halves.

Cryptic enhancer activity Sub-element pse_B5 appears to have stronger enhancer activity than that of the full 5' intergenic region of *D. pseudoobscura yellow* despite the fact that the spatial expression pattern it drives throughout each abdominal tergite is reminiscent of the pattern the intact 5' region drives. The spatial expression pattern that pse_B5 drives in the head and thorax is different than that of the full 5' intergenic region, however, where the strong expression pse_B5 sub-element drives encircles the eyes and ocellum in the head, and encompasses the length and width of the thorax, whereas the full 5' intergenic region drives expression only in a round spot on top of the head and in a weak manner in the thorax. This suggests that there may be “cryptic” enhancer activity in the 5’ intergenic region of *D. pseudoobscura yellow*, similar to what was observed in *D. melanogaster yellow* 5’ intergenic region.

Enhancer activities lying in the sub-elements of 5’ intergenic region of yellow *from* D. willistoni

Among the seven enhancer sub-elements of *D. willistoni yellow* 5’ intergenic region the activities for two of them remain to be determined. Another two do not appear to drive expression above background levels. Part of the expression pattern driven by one sub-element can be regarded as a fraction of the activity driven by full region, but the rest is cryptic, i.e., it is not a part of the expression pattern driven by the full region. The rest of the (two) sub-elements harbor cryptic enhancer activities only.

Minimal enhancer Among the five out of seven *D. willistoni* 5’ intergenic sub-elements tested, so far there is none that drives abdominal expression in a spatial pattern very similar to the pattern driven by the intact *D. willistoni* 5’ region, where expression is observed at the posterior end of each abdominal tergite as a horizontal stripe with a peak pointing towards the anterior in the middle.

Fraction Sub-elements will_C1 and will_C2 appear to drive expression throughout the length and width of each abdominal tergite, although it is notable that will_C1 seems to drive higher expression at the posterior halves of tergites A3, A4, A5 in females and A3 and A4 in males, which is reminiscent of the horizontal stripe observed in the expression pattern driven by the full *D. willistoni yellow* 5’ intergenic region.

Cryptic enhancer activity In females, will_C1 does not seem to drive expression above background levels in tergite A6, whereas in males tergite A5 and A6 show enhancer activity throughout their length and width, suggesting male-specific enhancer activity in the abdomen. It is harder to asses any sexual dimorphism in the expression driven by will_C2 due to weak activity, however in this transgenic line also males appear to show stronger expression in tergites A4, A5 and A6 as compared to females. Sub-element will_C4 also appears to show sexually dimorphic enhancer activity in the abdominal epidermal cells since males show significantly higher expression in tergites A4, A5 and A6 as compared to females, where enhancer activity appears close to absent in the epidermal or bristle cells. It is intriguing to find these sub-elements to show the above-mentioned activities since the full 5' intergenic region of *D. willistoni yellow* does not appear to have sexually dimorphic activity in the abdomen (Figure 3-2C8) indicating the existence of cryptic enhancer activity in this region. The thoracic expression observed in line will_C1 and the head expression observed in both will_C1 and will_C2 also are also cryptic since they do not appear in the expression driven by the intact 5' intergenic region of *D. willistoni yellow*.

Enhancer activities lying in the sub-elements of intron of yellow *from* D. melanogaster

Among the three sub-elements of *D. melanogaster yellow* intron one (mel_D2) remains to be tested. One (mel_D1) harbors a minimal enhancer as well as cryptic enhancer activity and one (mel_D3) does not seem to drive expression in epidermal or bristle associated cells above background levels.

Minimal enhancer mel_D1 drives expression in bristle associated cells throughout the abdomen and thorax in a spatial pattern highly similar to the one driven by the intact intron and hence can be regarded as a minimal bristle enhancer.

Cryptic enhancer activity Intriguingly, mel_D1 appears to drive expression also in the epidermal cells of the abdominal tergites A5 and A6 in males, but not females, suggesting sexually dimorphic activity. This is interesting because the full *D. melanogaster yellow* intron does not appear to drive any epidermal-cell expression, neither does its activity change between males and females. Hence the sexually dimorphic epidermal cell enhancer activity can be regarded as cryptic.

Enhancer activities lying in the sub-elements of intron of yellow *from* D. pseudoobscura

Among the four sub-elements from *D. pseudoobscura yellow* intron, enhancer activity of one (pse_E3) remains to be determined. Within the remaining three, one (pse_E1) does not appear to drive expression above background levels and one (pse_E4) appears to drive epidermal cell expression in the body in a spatial pattern highly reminiscent of the one driven by the full intron. The expression driven by the other sub-element (pse_E2) is in part a fraction of the total expression driven by the full intron, but part of it is cryptic.

Minimal enhancer pse_E4 appears to have enhancer activity highly reminiscent of that of the full intron such that both pse_E4 and the full *D. pseudoobscura yellow* intron drive expression in the epidermal cells in the thorax where there are eight fields of expression in a symmetrical pattern. Moreover, in the abdomen, both regions appear to drive expression in all tergites A2, A3, A4, A5, A6, more prominently on the anterior edge of each. The full intron, however, appears to drive stronger expression also laterally (on the two sides) in the male abdomen as compared to females, which is not observed in the expression pattern driven by pse_E4, making it an “incomplete” minimal enhancer.

Fraction pse_E2 drives sexually dimorphic expression, where in males abdominal tergites A5 and A6 show strong reporter gene expression throughout, and in A3 and A4 only laterally, but in females expression is not detectable in any of the abdominal tergites. This male-specific expression in the abdomen can be detected in the pattern driven by the full *D. pseuodoobscura yellow* intron as well, where males seem to show stronger reporter gene expression laterally in abdominal tergites A3, A4, A5 and and A6, as compared to females.

Cryptic enhancer activity pse_E2 also drives expression in the epidermal cells encircling the head, which is not observed as part of the expression driven by the full *D. pseudoobscura yellow* intron. This suggests existence of “cryptic” enhancer activity in the *D. pseudoobscura yellow* intron.

Enhancer activities lying in the sub-elements of intron of yellow from D. willistoni Among the four sub-elements of *D. willistoni yellow* intron, two remain to be tested for enhancer activity. One (will_F4) does not appear to drive expression above background levels in the body epidermal or bristle associated cells at the late pupal stage. One (will_F3) appears to harbor cryptic enhancer activity.

Cryptic enhancer activity will_F3 drives sexually dimorphic expression in the abdominal epidermal cells in a pattern highly similar to the one sub-elements pse_B2, will_C4 and pse_E2 drives, where tergites A4, A5 and A6 show strong expression in males as compared to the lack of expression in the whole abdomen in females. will_F3 also drives expression in the epidermal cells encircling the head in a pattern similar to the one observed from pse_E2. Neither of these two expression patterns are part of the one driven by the full *D. willistoni yellow* intron, suggesting cryptic enhancer activity lying in *D. willistoni yellow* intron.

Next I sought to identify the transcription factors that bind to *yellow* enhancer sub-elements in order to start understanding how these binding factors differ between different sub-elements.

Identifying sets of transcription factors that bind to *yellow* enhancer sub-elements

Previously, there were two transcription factors, Abdominal-B (Abd-B) and Engrailed (En), that were shown to bind *yellow* 5’ intergenic region in *D.*

melanogaster and *D. biarmipes*, respectively. In order to broaden the known set of transcription factors binding to *yellow* enhancers, I used a yeast-one-hybrid (Y1H) assay to screen *yellow* enhancer sub-elements from *D. melanogaster*, *D. pseudoobscura* and *D. willistoni* with a *D. melanogaster* transcription factor library, which is thought to harbor ~85% of all transcription factors in this species (Figure 3-3).

Briefly, using Y1H, I was able to screen 25 out of the 29 sub-elements from 5' intergenic and intronic regions of *yellow* from the three species with 647 *D. melanogaster* transcription factors (~85% of all *D. melanogaster* transcription factors). 19 out of the 25 sub-elements that were assayed with Y1H were also functionally tested for enhancer activity (Figure 3-5). Each of the 25 *yellow* enhancer sub-elements (the "bait") were put 5' of a reporter gene (His3) necessary for Histidine biosynthesis. Each transcription factor (the prey) was fused to a Gal4 activation domain (AD) to ensure activation of the reporter gene upon binding even if the transcription factor is a repressor. The bait and prey were put in the yeast cellular environment and the presence of interaction between a sub-element and a transcription factor was observed through the activation of the reporter gene, which is expected to happen through the direct binding of the transcription_factor-Gal4AD fusion protein to the bait DNA, however it may also happen indirectly and be mediated by an endogenous yeast protein. We were able to use two negative controls (no transcription factor-Gal4-AD construct and no transcription factor, but only Gal4-AD construct) to ensure that the endogenous yeast transcription factors or Gal-AD alone were not activating the reporter gene. Reporter gene activation was assayed on plates that contain increasing levels of 3-Amino-1,2,4-triazole (3-AT), a competitive inhibitor of the His3 enzyme. In order to be able to grow on increasing concentrations of 3-AT, transformants need to make higher amounts of the His3 enzyme than what is provided by basal level expression of *HIS3* gene.

For each *yellow* enhancer sub-element screened with Y1H, transcription factors showing interaction at only one level of 3-AT were marked as weaker interactors (Table 3-1, light blue shaded boxes) and transcription factors that showed interaction in more than one level of 3-AT were marked as stronger interactors (Table 3-1, dark blue shaded boxes). Overall, 204 out of the 647 *D. melanogaster* transcription factors tested were found to interact with at least one sub-element from the 5' intergenic or intronic enhancers of *yellow* from *D. melanogaster*, *D. pseudoobscura* or *D. willistoni* (Figures 3-4A and 3-4B). A total of 280 interaction events were detected since some transcription factors (41) were found to interact with more than one *yellow* enhancer sub-element screened (Table 3-1, Figure 3-4B). In total, the 5' intergenic regions of *yellow* from *D. melanogaster*, *D. pseudoobscura* and *D. willistoni* interacted with 154 transcription factors whereas the intronic regions interacted with 68. Separately, 73 transcription factors were found to interact with *D. melanogaster yellow* 5' intergenic region, 27 with *D. pseudoobscura* and 76 with *D. willistoni yellow* 5' intergenic regions. Moreover, 24 transcription factors were found to interact with *D. melanogaster yellow* intron, 21 with that of *D. pseudoobscura* and 27 with *D. willistoni yellow* intron. There were 18 transcription factors that were shared interactors among one or more 5' intergenic and intronic *yellow* enhancer sub-elements (Figure 3-4C).

Did Y1H identify the transcription factors that were previously shown to be real or candidate regulators of *yellow* expression?

Besides Abd-B and En, which are known to bind to *yellow* body and wing enhancers, respectively, there are few other transcription factors, which, based on prior knowledge, are good candidates as *trans* regulators of *yellow* expression. These transcription factors are Fruitless (Fru), Doublesex (Dsx), Bric-a-brac (Bab), Optomotor-blind (Omb) and T cell factor (TCF). Fru and TCF are known to be genetically upstream of *yellow* (Drapeau et al. 2003; Werner et al. 2010a) and Dsx, Bab, and Omb are previously shown regulators of pigmentation

patterns in flies (Kopp and Duncan 1997; Kopp et al. 2000) and as it is a pigmentation gene, they may be regulating *yellow* as well.
All of the above seven candidate transcription factors exists in the transcription factor library used in Y1H (Hens et al. 2011). Among them, only Abd-B, En, Dsx and Fru were found to interact with some of the *yellow* enhancer sub-elements tested. This may be due to the high false positive rate of Y1H (see Discussion), or because Omb, Bab and TCF are not direct regulators of *yellow* or, else, because these transcription factors are divergent enough in *D.pseudoobscura* and *D. willistoni* such that the *D. melanogaster* proteins cannot recognize their binding sites. Interestingly, unpublished data showed that Omb fails to bind *D. melanogaster yellow* 5' enhancers *in vitro* (personal communication, P.J.W.), suggesting that Omb may have been a true negative in the Y1H assay.

Abd-B interacted with sub-elements mel_A2, will_C2, and will_C4, Dsx was shown to interact with will_C4 and will_C5, Fru was shown to interact with will_C2 and En was shown to interact with will_C7. It is intriguing to find none of the candidate binders to interact with any of the *D. pseudoobscura* sub-elements even though some of them appear to drive male-specific expression pattern in abdominal tergites A5 and A6 (Figure 3-2B), which is one of the features of expression that Abd-B and Dsx have been shown to regulate. Moreover, Abd-B was able to bind *D. subobscura yellow* enhancer in vitro and *D. pseudoobscura yellow* 5' intergenic region carries the same binding site (Jeong et al. 2006a). This may be due to any of the three reasons stated in the previous paragraph. It is also interesting to find that all four of the above mentioned transcription factors were found to interact with *D. willistoni* sub-elements, not a species that was the subject of any of the studies involving Abd-B, Dsx, Fru or En. Last but not least, Abd-B and En were found to interact with 5' sub-elements, and not with any of the intronic elements, which is consistent with the previous studies showing they bind to 5' intergenic region of *yellow*.

Are enhancers with similar activities bound by similar sets of transcription factors?

Next, I sought to identify whether there is a correlation between the set of transcription factors that were shown to interact with a *yellow* enhancer sub-element and the enhancer activity of the particular sub-element. I grouped *yellow* enhancer sub-elements based on the similarity of the pool of transcription factors that were found to interact with them using the program Cluster (de Hoon et al. 2004) with default settings (Figure 3-4 - top tree). Complementarily, I grouped the transcription factors based on the similarity of the set of *yellow* enhancer sub-elements that they interact with (Figure 3-4 - left tree). Both cluster analyses were meaningful only for the 41 transcription factors that were found to interact with more than one *yellow* enhancer sub-element and hence were conducted only for those. Subsequently, enhancer activities of the *yellow* sub-elements were matched with their transcription factor binding profiles (Figure 3-4). A qualitative assessment of the similarities and differences between the expression patterns driven by different *yellow* sub-elements at the late pupal stage were determined based on the spatial pattern and strength of expression in the epidermal as well as bristle cells in the abdomen, thorax and head as well as based on whether the enhancer activity is male-specific in abdominal tergites A5 and A6.

The results can be described in three categories:
1- *yellow* enhancer sub-elements with similar transcription factor binding profiles but different enhancer activities
2- *yellow* enhancer sub-elements with similar transcription factor binding profiles and similar enhancer activities
3- *yellow* enhancer sub-elements with different transcription factor binding profiles, but similar enhancer activities

1- yellow *enhancer sub-elements with similar transcription factor binding profiles but different enhancer activities*

The two *yellow* enhancer sub-elements that have the most similar transcription factor binding profiles are 5’ intergenic pse_B6 and intronic pse_E4, however since enhancer activity of pse_B6 is not yet determined, one cannot evaluate how the transcription factor binding profiles of the two sub-elements relate to their enhancer activities. Transcription factor binding profiles of pse_B6 and pse_E4 appear to have equal level of similarity to that both of pse_B5 and pse_E1, where the latter two also share the same level of similarity. This is the second highest similarity in transcription factor binding profiles observed between *yellow* enhancer sub-elements. It is intriguing that all four of these sub-elements belong to *D. pseudoobscura*.

Additionally, both pse_B5 and pse_E4 drive expression in the epidermal cells in the abdomen and thorax, where there are both similarities and differences in spatial patterns. (For instance, both sub-elements drive expression through the length and width of each abdominal tergite, and throughout most of the thorax. However, pse_B5 appears to drive equally strong expression thorughout abdomen and thorax and encircling the head, but pse_E4 appears to drive stronger expression not throughout each abdominal tergite, but mainly in the circumference of each abdominal tergite. Also this sub-element does not drive expression throughout the thorax, but in a symmetrical pattern that encompasses 8 separate rectangular patches, and seems to lack enhancer activity driving expression in the head.) Moreover the strength of expression coming from pse_B5 appears stronger than the one coming from pse_E4. The difference in enhancer activity is even more drastic when pse_E1 is compared to pse_B5 and pse_E4, since the former sub-element does not seem to drive visible expression at the late pupal stage in any of the cell types investigated. This suggests that even though the transcription factor binding profiles of the three *D. pseuodoobscura yellow* enhancer sub-elements showed higher similarity among each other as compared to the rest of the *yellow* enhancer sub-elements that did not group with them, differences between the pools and the binding composition of the transcription factors that bind to pse_B5, pse_E4 and pse_E1 were sufficient to alter enhancer activity.

The level of similarity between the transcription factor binding profiles of pse_B4, will_C7 and will_F2 are close to the level of similarity observed among the above-mentioned four *D. pseudoobscura yellow* enhancer sub-elements. However, the enhancer activities of these three sub-elements are not yet identified hence we cannot interpret how the similarity in transcription factor binding profiles correlates to enhancer activities. Interestingly these three sub-elements, as well as mel_A1, will_C4 and mel_D2 overall appear to be more similar to each other (Figure 3-4, blue rectangle), based on their transcription factor binding profiles, than they do to the rest of the sub-elements. This appears to be mostly due to one shared transcription factor, Neurofibromin 1 (Nf1) among these six sub-elements.

2- yellow *enhancer sub-elements with similar transcription factor binding profiles and similar enhancer activities*

The most intriguing result of this analysis is the similarity in the transcription factor profiles of five *yellow* enhancer sub-elements that drive sexually dimorphic expression in the abdomen, mel_A2, mel_A4, will_C2, pse_B1 and will_C1, with respect to their proximity to each other in the tree (Figure 3-4, red rectangle). All five of these sub-elements appear to drive stronger expression in one or more of the abdominal tergites A4, A5 and A6 in males as compared to females. Even though sexual dimorphism is a shared feature of the expression patterns driven by the five sub-elements, these enhancer activities differ in other spatial patterns of expression. For instance, between mel_A2 and will_C2, both of the sub-elements drive expression in the epidermal cells in the abdomen and head, whereas only mel_A2 appears to drive expression in the epidermal cells of the thorax. The spatial pattern of expression driven by the two sub-elements seem to differ both in the abdomen and head, where mel_A2 drives expression at the posterior end of each abdominal tergite and both on top of the head and around the eyes, whereas will_C2 drives expression throughout the length and width of each abdominal tergite and only on top of the head but not around the eyes.

3- yellow *enhancer sub-elements with different transcription factor binding profiles, but similar enhancer activities*

Two intronic sub-elements, mel_D1 and will_F3, show male-specific enhancer activity in the abdomen, but share only one transcription factor, Suppressor of variegation 3-7 (Su(var)3-7), which has not previously shown to be expressed in a sexually dimorphic manner. In my Y1H analysis Su(var)3-7 was found to also interact with sub-elements that do not appear to have sexually dimorphic activity and hence may be necessary for general epidermal-cell activity.

pse_B2 and will_C4, which also show same type of sexually dimorphic expression in abdominal tergites (part of) A4, A5 and A6, did not group with the rest of the sexually dimorphic sub-elements based on transcription factor binding profile similarity. They appear to share some transcription factors with some of the other sexually dimorphic enhancer-sub-elements. For instance, pse_B2 shares two transcription factors, Jun-related antigen (Jra) and Nubbin (nub) with two other sexually dimorphic elements, pse_B1 and will_C2, respectively. Since pse_B1 and pse_B2 are neighboring overlapping sub-elements from *D. pseudoobscura yellow* 5' intergenic region, Jra may be binding to the region overlapping between the two sub-elements. pse_B2 and will_C2 both have epidermal cell enhancer activity, for which Nub may be important for. Moreover, these two sub-elements correspond to neighboring sub-elements, but in different species. Hence the two sub-elements may be sharing ancestral binding sites for transcription factors, including Nub and others that Y1H missed.

Other sub-elements that showed highly similar enhancer activities also failed to show similarity in their transcription factor binding profiles. For instance, mel_A3 and pse_B5 both show strong expression in the epidermal cells throughout the abdomen and thorax and encircling the head, however their transcription factor binding profiles look highly different in that they do not seem to share any binding factors. These two sub-elements are particularly interesting because they are both located in the 5’ intergenic region of *yellow* and are similar in their proximity to transcription start site.

Evaluating the *in vivo* phenotypic effects of some of the transcription factors identified in Y1H as possible interactors of *yellow* enhancers

Y1H is a hypothesis generating technique rather than a definitive one. Because of this, it is important to test Y1H results using an independent technique. One of these approaches is to test whether the transcription factors that were shown to interact with a certain enhancer element using Y1H also do so in the cellular environment of the organism they come from. An indirect way of testing this is to knock down a candidate transcription factor and look for phenotypic effects related to the putatively regulated gene.

I conducted an experiment to look at changes in pigmentation caused by knocking down transcription factors that were shown to interact with *yellow* enhancer sub-elements from *D. melanogaster*, *D. pseudoobscura* and *D. willistoni*. Pigmentation is an appropriate phenotype to look at since *yellow* gene is required for black pigment formation in the body, wings and bristles in adult flies and it is straightforward to score. I used the *pannier(pnr)*-Gal4 driver with 43 UAS-RNAi lines (Table 3-3) from the TRIP (Transgenic RNAi Project) collection against 40 transcription factors (out of 204 total) identified as a result of Y1H. Among these 41 transcription factors, 29 were found to interact with more than one *yellow* enhancer sub-element whereas the rest (12) appeared to interact with a single *yellow* enhancer sub-element (Table 3-1, transcription factors highlighted in yellow). In adult flies *pnr* drives expression in the dorsal midline along the dorsal anterior-posterior length of the flies ((Heitzler et al. 1996): Figure 3-4; (Wittkopp et al. 2002a): Figure 3-3D), which is very useful for looking at effects on pigmentation in body epidermal and bristle cells (though not wing cells) since that is the most highly pigmented part of the adult fly. As a result, any phenotypic effects of transcription factor knockdowns should be observed in the dorsal midline, whereas the part of the abdomen, thorax and head that are left outside of the dorsal midline, i.e., outside the *pnr*-expression domain, should have wild-type pigmentation, which can be used as an internal control.

As a result of knocking down 41 different transcription factor genes using *pnr*-Gal4 driver, I did not observe lethality in the progeny of any of the crosses, but I did find that the line harboring a UAS-RNAi transgene for the transcription factor Jra, when activated by *pnr* driver, gave rise to females only. I found four transcription factors which upon getting knocked down affected pigmentation in the abdomen or thorax. These four transcription factors are Abd-B, Dsx, Fru and Ventral veins lacking (Vvl) (Figure 3-5). Among these four, Abd-B and Dsx were previously shown to affect abdominal pigmentation, whereas the effects of Fru and Vvl on pigmentation are documented for the first time. Below I describe the observed pigmentation phenotypes after knocking down the above-mentioned four transcription factors.

Abd_B As a result of Y1H, this transcription factor was found to interact with sub-elements mel_A2, will_C2 and will_C4. When knocked down with RNAi in *D. melanogaster*, male flies carrying *pnr->abd-B*-RNAi showed lack of male-specific black pigment in the dorsal midline of abdominal tergite A5. Hence in this region pigmentation looked similar to that of females. Abdominal pigmentation in female flies carrying *pnr->abd-B*-RNAi did not seem to be affected (Figure 3-5). This is consistent with previous studies showing that loss of function mutation in *abd-B* leads to loss of male-specific pigmentation in the abdomen (tergites A5 and A6) (Celniker et al. 1990; Hopmann et al. 1995), but the phenotypic effect of knocking down Abd-B is not as strong as the effect of its knock-out mutant.

Dsx As a result of Y1H, this transcription factor was found to interact with only *D. willistoni* 5' intergenic *yellow* enhancer sub-elements, will_C4 and will_C5. When knocked down with RNAi in *D. melanogaster*, female flies carrying *pnr->dsx*-RNAi showed male-specific black pigment formation in the dorsal midline of abdominal tergite A6. Males carrying *pnr->dsx*-RNAi did not seem to have altered body pigmentation (Figure 3-5). This is consistent with studies showing that loss of function in *dsx* cause ectopic pigmentation in female abdominal tergite A5 and A6 (Baker and Ridge 1980; Couderc et al. 2002), but similar to the case in Abd-B, the phenotypic effect of knocking down Dsx is not as strong as the effect of its knock-out mutant.

Fru Y1H showed that this transcription factor interacted with sub-element will_C2 only. When knocked down with RNAi in *D. melanogaster* flies, females carrying *pnr->fru*-RNAi appeared to have decreased pigmentation in the dorsal midline of abdominal tergites A5 and A6, however in males the phenotype appeared to be weaker such that only a thin line (red arrow) in the middle of abdominal tergite A5 had faded pigmentation. Fru is known to be genetically upstream of *yellow* affecting its function on male mating behavior (Drapeau et al. 2005). However, effects of Fru on pigmentation is identified for the first time.

Vvl As a result of Y1H, this transcription factor was also found to interact with will_C2 only. When knocked down by RNAi in *D. melanogaster*, both male and female flies carrying *pnr->vvl*-RNAi showed a very clear phenotype where in the dorsal midline of the adult cuticle black pigment was virtually lacking in all abdominal tergites whereas the two sides of the dorsal midline had wild type pigmentation (Figure 3-5). Effects of Vvl on *yellow* regulation or pigmentation has never been previously documented.

It is important to note that looking at the effects of knocked-down transcription factors, which were shown to interact with *yellow* enhancer sub-elements, on pigmentation, particularly black pigment formation, has limitations in validating

Y1H results. Any effects seen on pigmentation upon knocking down a transcription factor does not validate direct binding of the particular transcription factor with the *yellow cis*-regulatory region. There may be intermediate genes/ proteins between the particular transcription factor and *yellow*. Moreover, the knocked-down transcription factor may be showing its effect on pigmentation by affecting another pigmentation gene (such as *tan* that is required for brown pigment formation) without affecting *yellow* expression at all. This is not known since expression of *yellow* or other pigmentation genes were not measured in this study. However one may be able to sort this out by specifically looking at reduction or increase in black pigment formation or better, in future work, reduction or increase in GFP activity in the relevant reporter genes.

The rest of the 37 transcription factors tested did not seem to affect pigmentation in adults when knocked-down (Figure 3-6). For example, Figure 3-6 shows knock-down phenotypes for transcription factors Hr78 and Hr38, both of which appeared to interact with one or more yellow enhancer sub-elements from the *yellow* 5' intergenic region, and not the intron of all three species. Hr78 was found to interact with 7 sub-elements in total, whereas Hr38 interacted with 6. Neither of the two transcription factors appeared to alter pigmentation in the dorsal midline upon getting knocked-down. This may be due to the fact that they are false positives and do not actually interact with *yellow cis*-regulatory elements. However, they may be real direct regulators of *yellow*, but could be missed by this type of RNAi study because the RNAi constructs may not have effectively knocked down the transcription factors, or these transcription factors may be real regulators of *yellow* at developmental stages that were not investigated in this study (for instance during larval stage, where *yellow* is necessary for black pigment formation in larval mouthparts), else knocking down only one transcription factor may not be sufficient to change *yellow* expression strongly enough to affect pigmentation.

Discussion

Overall, using *D. melanogaster* as the common transgenic host, I identified the enhancer activities lying in 21 out of 29 1-kb sub-elements spanning 5' intergenic and intronic regions of *yellow* from three Sophophora group species, *D. melanogaster*, *D. pseudoobscura* and *D. willistoni*. As a result, I found that most (15/21) of the tested sub-elements of *yellow* 5' intergenic and intronic regions harbor enhancer activities that drive expression in body epidermal cells in the late pupal stage. Among these, as compared to the total expression pattern driven by the full region that they are isolated from, some sub-elements drive expression that almost recapitulates the total expression pattern (minimal enhancer), some drive only a fraction of the total expression pattern and some drive expression in spatial patterns that are not part of the total expression pattern (cryptic enhancer).

A yeast-one-hybrid screen (Y1H), where the above-mentioned enhancer sub-elements were assayed with a *D. melanogaster* transcription factor library, which harbored ~85% of all transcription factors in this species, identified a large set of candidate transcription factors binding to *yellow* 5' intergenic and intronic regions from the three Sophophora group fly species. As was already indicated by a correlation between the expression pattern driven by *D. pseudoobscura* and *D. willistoni yellow* enhancers in the common host *D. melanogaster* and the pigmentation patterns of the donor species, Y1H also showed that the *trans* environment appeared to be conserved between *D. melanogaster* and the other two species. I found that similar number of transcription factors interacted with *D. willistoni yellow* 5' intergenic and *D. pseudoobscura* and *D. willistoni* intronic regions as did with *D. melanogaster* ones, suggesting that the transcription factor binding sites are conserved between these three species. Only ~10-20% of the transcription factors were shared binders between physically homologous regions from different species even though they drive similar expression. This may be due to the high false negative rate of Y1H. In fact RNAi validation showed that some of the transcription factors that were shown to bind to *D. willistoni yellow* enhancers only, have effects on pigmentation in *D. melanogaster* suggesting that these proteins control *yellow* expression *D. melanogaster* as well. Overall, RNAi validation identified two transcription factors (Abd-B and Dsx) that were previously shown to affect pigmentation in Drosophila, but it also identified two other transcription factors (Fru and Vvl), whose effects on pigmentation are novel.

Enhancer activities of *D. melanogaster yellow* 5' intergenic sub-elements are consistent with previous functional analyses

D. melanogaster yellow enhancers were first identified by Geyer et al. (Geyer and Corces 1987b) and Martin et al., (Martin et al. 1989a) using pigmentation in the abdomen, wings and bristles as a readout of *yellow* expression. As a result, it was shown that the *yellow* 5' intergenic region is necessary for the gene's expression in the body and wings whereas the intron is necessary for bristle expression. Subsequent studies characterized the enhancers, especially in the 5' intergenic region of *yellow* from *D. melanogaster* (Wittkopp et al. 2002c; Drapeau et al. 2003; Jeong et al. 2006a), Chapter 2) as well as from other species (Wittkopp et al. 2002c; Jeong et al. 2006a), Chapter 2). Among these studies Jeong and colleagues (Jeong et al. 2006a) dissected the 5' intergenic region of *D. melanogaster yellow* in detail and the results of my study are consistent with their results.

First of all, Jeong and colleagues did not look beyond -2869 relative to the transcription start site, which corresponds to the 5' end of mel_A2. However, in my study the enhancer activities up to -3815 were identified, hence the enhancer activity of mel_A1 was described for the first time here. Despite the lack of mel_A1, the full 5' intergenic region Jeong and colleagues looked at drove virtually identical expression to the full 5' intergenic region investigated in my study. This suggests that the activity of mel_A1 is not necessary for the full expression pattern driven by *D. melanogaster yellow* 5’ intergenic region. They found that a 5’ intergenic element extending to position -1867, which, in my study, corresponds to sum of sub-elements mel_A3, mel_A4 and mel_A5, drives expression through the length and width of all abdominal tergites with increased strength in tergites A5 and A6 in males as compared to females. This locates the “stripy” expression pattern, i.e., expression only in the posterior half of each abdominal tergite with a middle peak pointing towards the anterior, to between -2869 and -1867, which approximately corresponds to sub-element mel_A2, consistent with the identified enhancer activity of this region in my study. Lastly, Jeong and colleagues localized a sexually dimorphic enhancer activity to where mel-A4 corresponds to, again consistent with my study. The authors also found, through in vitro EMSA and in vivo binding site deletion assay, that Abd-B directly binds to mel_A4 region and contributes to the sexual dimorphic expression that this region drives in the abdomen. Interestingly, the Y1H experiment I conducted did not confirm this, perhaps due to high false negative rate of the technique. I found, however, that Abd-B binds to mel_A2, which was not widely investigated by Jeong et al. (2006). It is likely that Abd-B is a major regulator of male-specific abdominal expression (tergites A5 and A6) driven by both mel_A2 and mel_A4.

Possible reasons for observing cryptic enhancer activity in the sub-elements of *yellow* 5’ intergenic and intronic regions

Besides the sub-elements that recapitulate (almost) all (minimal enhancer) or a fraction of the total enhancer activity of the full region they are isolated from, it was particularly intriguing to observe sub-elements that drove expression in patterns that were not part of the expression pattern driven by the full region (cryptic enhancer activity). This may have several explanations:

1- Change in the proximity of a sub-element to the transcription start site can alter its enhancer activity.

Enhancer elements typically act independent of their proximity to the transcription start site. However, we know that for certain enhancer elements this proximity is important for their function, such that decreased proximity results in elevated levels of expression (Swanson et al. 2010). Among the *yellow* enhancer sub-elements that show cryptic enhancer activities, this may be particularly true for the ones that drive expression in spatial patterns that are similar to but stronger than the expression pattern driven by the full region. The thoracic and abdominal expression driven by sub-element pse_B5 sets a good example for this possibility. (It is worth noting however, that the spatial pattern of pse_B5 driven expression is not fully identical to the one full *D. pseudoobscura* 5' intergenic region drives, especially in the head.) One can assess if this possibility is real by testing the particular sub-element for enhancer activity when located equally distant from the transcription start site of the reporter gene as it is from the *yellow* transcription start site.

2- The altered genomic position of the sub-element from the endogenous position can alter the expression pattern it drives.

It is possible that the foreign genomic locus on the 2nd chromosome, where the reporter gene constructs are integrated, is adjacent to sequences that harbor information that drives expression in the epidermal cells in the body at the late pupal stage, and that is why some sub-elements drive expression in spatial patterns that are not part of the expression pattern driven by the full region. The fact that some sub-elements failed to drive expression above background levels and the full region was tested in the same genomic location complicates this possibility such that the sequences flanking the integration site activate expression in body epidermal cells only in combination with the appropriate sequences within a sub-element (and they don't have the same combinatorial effect with other sub-elements or the full region). Testing the activity of *yellow* enhancer sub-elements in other integration sites would assess the veracity of this possibility.

3- The cryptic activities of yellow *enhancer sub-elements may be repressed in their native loci and be revealed when they are tested in isolation.*
It is possible that part or all of the expression pattern driven by *yellow* enhancer sub-elements with cryptic activity, when they are located in their native loci, is repressed or simply altered by the surrounding sequences. The break points of sub-elements may separate certain transcription factor binding sites from each other, which, when together drive expression that is part of the total expression pattern driven by the full region.

This third possibility is particularly intriguing because recent examples show how existing enhancer activities in a region can accelerate change in gene expression patterns. In one example a latent enhancer activity is repressed due to few point mutations and revealed as a result of changes in these repressive sequences (Prabhakar et al. 2008; Sumiyama and Saitou 2011). In another case an enhancer activity started out weak and over time became fully repressed or stronger due to repressing or activating mutations, respectively (Rebeiz et al. 2011a).

Changing latent enhancer activities may be a common trajectory especially for rapidly changing enhancers. Despite the fact that transcription factor binding sites can be created rapidly, relative to the evolutionary distances between species (Stone and Wray 2001), arguably fewer mutations are needed for an enhancer activity to come about in a region that is already primed for it (i.e., if it already harbors part of the transcription factor binding sites necessary to drive a certain expression pattern) than the ones needed to create binding sites from putatively non-functional sequence (i.e., from scratch). This is because there is a particular set of transcription factors that are expressed in the right tissue and time to affect a cognate gene’s expression, and the number of mutational events necessary to create all the appropriate transcription factor binding sites from scratch would be a multiplicate of the few mutations needed to add onto an existing group of appropriate transcription factors.

If one can eliminate the first two possibilities with appropriate experimentation, the third possibility, existence of latent activities in the 5' intergenic and intronic regions of *yellow*, can explain the dynamic architecture and activity of *yellow* enhancers.

A large set of candidate transcription factors regulating *yellow* expression was identified

Despite its widely used mutant form and relatively well studied *cis*-regulatory elements, until this study there was limited knowledge on the transcription factors regulating *yellow* expression. The conducted Y1H experiment found that 204 out of the 647 transcription factors from the *D. melanogaster* library bind to at least one enhancer sub-element tested. This is perhaps not all of the transcription factors that bind to *yellow* enhancer sub-elements tested since Y1H is known to have a high false negative rate (~74%, (Hens et al. 2011)). For instance, among the 154 transcription factors that were found to bind to 5' intergenic regions of *yellow* from *D. melanogaster*, *D. pseudoobscura* and *D. willistoni,* only 19 (~12%) were shared between species, which is arguably less than expected given the similarities in the expression these three regions drive, i.e., same cell type, developmental stage, etc. This was true for introns as well, among the total of 68 transcription factors found to interact with *yellow* introns from *D. melanogaster*, *D. pseudoobscura* and *D. willistoni,* only 3 (~4.5%) were shared between species. This percentage is even lower than that of shared binding factors for 5' intergenic region partly because of the 4 (out of 11) intronic sub-elements that were not assayed with Y1H.

It is important to also note that the false positive rate of Y1H has been estimated to be ~23% (Hens et al. 2011). Hence not all of the 204 transcription factors identified are real binders of *yellow* enhancers. Further validation assays can eliminate these false positives to reveal a more accurate set of transcription factors controlling *yellow* expression.

Transcription factors that were found to bind more than one yellow *enhancer sub-element*

From the 204, 163 transcription factors interacted with a unique enhancer sub-element, and 41 interacted with more than one and up to eight sub-elements. These “multiple binders” are of particular interest because they may be responsible for some of the commonalities observed between the expression patterns of 5’ intergenic or intronic regions of *yellow* among Drosophila species. So far Y1H identified Neurofibromin 1 (Nf1) to bind to eight *yellow* enhancer sub-elements, Hormone-receptor-like-in-78 (Hr78) and Suppressor of variegation 3-7 (Su(var)3-7) appear to bind 7, Hormone-receptor-like-in-38(HR38) binds to 6, and CG5591, Ecdysone-induced protein 78C (Eip78C), sequoia (seq) and Regulatory factor X (Rfx) appear to bind 4 different enhancer sub-elements (Figure 3-4). These transcription factors have not previously been implicated in regulation of *yellow* or pathways that *yellow* is known to be involved in (pigmentation and male mating behavior). However, knowing what these proteins do can aid a better understanding of *yellow* regulation as well as function. To set a starting point to investigating the *trans* regulators of *yellow* and the biological pathways the gene may be involved in, below I give descriptions of the above mentioned “multiple binders” and their possible associations to *yellow* regulation:

An important tumor suppressor gene as the regulator of yellow:

Nf1 This protein was found to bind to at least one sub-element from all *yellow* 5’ intergenic and intronic regions dissected. Namely mel_A1, pse_B4, will_C2, will_C4, will_C7, mel_D2, pse_E1, and will_F2, were found to interact with Nf1.

Nf1 is a known tumor suppressor gene that encodes for a highly conserved GTPase activating domain as well as a C-terminal domain that regulates cyclicAMP levels. In humans null mutations are responsible for Neurofibromatosis type 1 disease which leads to learning defects, small stature, tumors and skin pigmentation abnormalities (Lee and Stephenson 2007). Null mutations of *Nf1* in flies are not lethal but lead to overall reduction in body size throughout all postembryonic stages (The et al. 1997). These mutants also show defects in learning and circadian rest-activity rhythm, loss in short and long-term memory (Guo et al. 2000; Ho et al. 2007) as well as increased sensitivity to oxidative stress, and shortened life spans (Tong et al. 2007). Overexpression of the gene leads to increased life span, improved reproductive success and increased resistance to oxidative stress. Except for improved reproductive success, none of the above mentioned functions of *Nf1* appear to be directly related to *yellow*'s known roles in pigmentation and behavior (male mating success). However given that Nf1 binds to a relatively high number of yellow enhancer sub-elements and is expressed throughout fly development, perhaps it is a major regulator of *yellow* and *yellow* is involved in some of the pathways that *Nf1* is involved. If in vitro and in vivo direct binding and functional assays confirm Nf1 as a direct regulator of *yellow*, this would open a new area of research looking at new roles of *yellow* in fly development and physiology.

Regulation of yellow *by the ecdysone pathway:*

Hr78 This transcription factor was found to bind to only yellow 5‘ intergenic sub-elements from all three species, namely mel_A2, mel_A3, mel_A4, mel_A5, pse_B1, will_C1 and will_C5. This gene is a target of the only steroid hormone in flies, ecdysone (also known as 20-hydroxyecdysone - 20E), which is a critical temporal signal for the major postembryonic developmental transitions in flies, such as molting of the larval cuticle and puparium formation. Hr78 is expressed throughout development with peaks during third instar larva and prepupal stages, which correlate with ecdysone pulses in the animal. The expression peaks of *yellow* (16-24hr embryo, L2-12hr L3 larva and 2-3 day post puparium formation) and *Hr78* do not perfectly match, but Hr78 is expressed in high amounts during all yellow peak expression times. Its precise function is not known, but null mutations of *Hr78* lead to lethality during third instar larval stage and polytene chromosome studies during the prepupal stage showed the colocalization of the protein with known ecdysone regulated puff loci. None of these loci correspond to that of *yellow*'s, but the authors looked at only previously identified ecdysone regulated loci (Fisk and Thummel 1998). Intriguingly, in the butterfly species *Papilio xuthus*, 20E was found to promote yellow expression and affect black pigment formation in the cuticle (Futahashi and Fujiwara 2007) suggesting the ecdysone pathway may be partly responsible for regulation of *yellow* in pigmentation in Drosophila as well. This suggests that some of the ecdysone target genes may regulate *yellow* for its role in pigmentation.

Hr38 This transcription factor was found to bind to sub-elements of yellow 5' intergenic regions, but no intronic regions, from all three Drosophila species. Namely these sub-elements are mel_A2, mel_A4, pse_B6, will_C2, will_C3, will_C5. Similar to Hr78, Hr38 is also downstream of the ecdysone pathway. Specifically, a null mutation in the gene leads to lethality as a result of reduced cuticle gene expression and consequent loss of cuticular integrity (ruptured cuticle) at the late pupal stage (Kozlova et al. 2009). It is expressed at low levels throughout development, but, similar to *yellow*, appears to peak during the late pupal stages (Kozlova et al. 2009). *yellow* is not included in the group of cuticle genes investigated by Kozlova and colleagues (Kozlova et al. 2009) since it is not known to have a role in cuticle formation, but rather it is important for the pigmentation of this structure. Regardless, given that the developmental time and tissue of expression of the cuticle genes are similar to that of *yellow*, *yellow* may also be regulated by the transcription factors controlling cuticle formation genes. In addition to its role in cuticle formation, *Hr38* is also involved in muscle carbohydrate homeostasis. Moreover, it is also expressed at high levels in adult brain and regulates *DOPA decarboxylase* (*Ddc*) expression suggesting also a neurological role for *Hr38*. Interestingly, both *yellow* and *Ddc* have major roles in pigmentation (Wittkopp et al. 2002a) in addition to their neurological functions, and now my data shows that they also share at least one regulator.

Eip78C This transcription factor was found to bind to sub-elements from *D. melanogaster* and *D. pseudoobscura* yellow 5' intergenic and intronic regions. Namely these sub-elements are mel_A2, pse_B6, mel_D3, pse_E4. *Eip78C* is also a target of the ecdysone pathway, but not as widely studied as *Hr78* and *Hr38*. It is expressed in the embryo, larva and pupa. Its expression peaks during the late pupal stages, which parallels *yellow* expression. Absence of *Eip78C* did not cause any abnormalities under laboratory conditions (Russell et al. 1996), which suggests that it is not a master regulator, but perhaps is responsible for fine tuned regulation of other ecdysone target genes.

Regulation of yellow *at the chromosomal level*:

Su(var)3-7 This protein was found to bind to sub-elements from all yellow 5' intergenic and intronic regions dissected except for *D. melanogaster* 5' intergenic region. Namely these sub-elements are, pse_B5, pse_B6, will_C1, will_C7, mel_D1, pse_E1, pse_E4, and will_F3. Su(var)3-7 is known to be mainly associated with (pericentromeric) heterochromatin and telomeres (but also some euchromatic sites) (Reuter et al. 1990). It is a modifier of position effect variegation (PEV) (gene silencing induced by heterochromatin), and changing doses of Su(var)3-7 can enhance or suppress PEV (Cléard and Spierer 2001). It has a particular role on X chromosome morphology and expression in males, such that reduced levels of Su(var)3-7 leads to a bloated X chromosome, whereas increased levels of Su(var)3-7 cause highly condensed (and silenced) X chromosome. This is thought to be because changes in Su(var)3-7 levels disrupt the correct distribution of the dosage compensation complex in males and lack of Su(var)3-7 cause male lethality (Spierer et al. 2008). Based on this information Su(var)3-7 appears to be a chromosomal level transcriptional regulator of, perhaps all, genes on the X chromosome. Being located on the X, *yellow* may be one of the targets of Su(var)3-7, but this *trans* regulator does not appear to be specific to *yellow* only.

Regulation of yellow *in bristle associated cells*:

seq This transcription factor was found to bind to three 5' intergenic sub-elements mel_A2, mel_A5, will_C2 and one from *D. willistoni* yellow intron, will_F2. Seq is expressed throughout fly development, but most highly during embryonic stages. It is crucial for normal morphogenesis (of axons and dendrites) of almost all neuronal types studied (Brenman et al. 2001). It has also been found to be important for external sensory organ (bristle) development, such that absence of Seq leads to loss of bristles (and gain of extra socket cells, one of the five types of bristle associated cells) (Andrews et al. 2009). We know that *yellow* is required for bristle pigmentation and hence can be controlled by transcription factors expressed in bristle associated cells, such as Sequoia. Notably, however, Y1H showed that Seq does not only bind to intronic regions that harbor the *yellow* bristle enhancer, but it also appears to bind 5' intergenic regions that harbor epidermal cell enhancer activity as well as necessary regulatory sequences for male mating success, suggesting that the regulatory role of this transcription factor is not restricted to bristle associated cells.

Rfx This transcription factor was found to bind to 5' intergenic sub-elements mel_A2, mel_A4, wil_C2 and will_C3. It is expressed in the peripheral nervous system and brain throughout Drosophila development and in testis in adults (Durand et al. 2000; Vandaele et al. 2001). It is also expressed in sensory organ precursor cells after puparium formation. Rfx, in Drosophila, is identified as an essential regulator of ciliated sensory neuron differentiation and lack of it leads to defects in chemosensory and mechanosensory behaviors during larva and adult stages. These defects are in part due to aberrations in mechanosensory bristle electrophysiology. This suggests that, similar to Seq, Rfx may also be a regulator of *yellow* expression in bristle associated cells (Dubruille et al. 2002) even though Y1H did not find it to bind any intronic sub-elements, perhaps an effect of the false negative rate of Y1H. The fact that Rfx was found to bind *yellow* 5' intergenic regions, suggests that its regulatory role on *yellow* is not restricted to bristle associated cells.

Transcription factors with minimal knowledge known about them:
CG5591 This transcription factor was found to bind to sub-elements from *D. melanogaster* and *D. pseudoobscura yellow* 5' intergenic regions only. Namely, these sub-elements are mel_A2, mel_A4, mel_A5, pse_B1. There is limited knowledge about this particular transcription factor. It was found to be a muscle specific lipid regulator and was implemented as one of the genes that may affect metabolic disorders (Pospisilik et al. 2010). Given the relatively high number *yellow* enhancer sub-elements it was found to interact with, this transcription factor remains as a strong candidate regulator of *yellow* expression, and further studies can show which function of *yellow* CG5591 is important for. We have limited information on many more candidate transcription factors that were found by Y1H to bind *yellow* enhancers. Further functional studies can shed light onto both the physiological and regulatory roles of these previously unstudied transcription factors.

New and old interactors of *yellow*

I was able to test 29 out of the 41 "multiple binder" transcription factors (Table 3-1), including Nf1, Hr78, Hr38 and Rfx. Interestingly, only few of the multiple binders showed an effect on pigmentation (Abd-B and Dsx), but none of the above strong candidates showed a phenotypic effect on pigmentation upon knock-down. This was perhaps because some RNAi lines used were not effectively knocking down the target gene, or perhaps these transcription factors do not control *yellow* expression at pigmentation related tissues and developmental time, or perhaps in some cases knocking down one transcription factor did not alter *yellow* expression enough to affect pigmentation. It is also possible that some of the multiple binder transcription factors were false positives, but given the 23% estimate of the false positive rate of Y1H, it is plausible to think that most are true interactors of *yellow* enhancers. Further experiments testing direct binding of these transcription factors to *yellow cis*-regulatory regions both *in vitro* and *in vivo* and their effects on *yellow* expression can shed light onto which of the above possibilities are true.

Regardless of whether they were shown to bind to multiple or single sub-elements of *yellow*, the four transcription factors (Abd-B, Dsx, Fru, Vvl) that altered adult pigmentation in *D. melanogaster* upon knock down are intriguing.

Among the four transcription factors that showed an effect on pigmentation upon knock-down, Abd-B and Dsx have previously been shown to affect abdominal pigmentation in a sexually dimorphic manner and their knockdown phenotypes confirmed that result (lack of male-specific pigmentation in abdominal tergites A5 and A6 in *pnr*->*abd-B*-RNAi, and increase in pigmentation domain in A6 and part of A5 in females harboring *pnr*->*dsx*-RNAi).

It is known that Abd-B shows its effect on pigmentation by directly regulating *yellow*. *In vitro* binding and *in vivo* functional tests showed that this transcription factor binds to sub-element mel_A4 (Jeong et al. 2006a). However, Y1H conducted in my study found Abd-B to bind mel_A2, but not mel_A4. Additionally, even though *in vitro* binding of Abd-B to *D. subobscura yellow* 5' intergenic region was previously shown (Jeong et al. 2006a), Y1H did not find Abd-B binding the yellow 5' intergenic region of *yellow* from the closely related species *D. pseudoobscura*. These seemingly missed binding events are perhaps a result of the high false negative rate of Y1H. As a novel binding event, however, Y1H did show that Abd-B binds to two sub-elements of *D. willistoni yellow* 5' intergenic region (will_C2 and will_C4). Functionally, in *D. melanogaster*, Abd-B is necessary for the male-specific expression of *yellow* in abdominal tergites A5 and A6, which is necessary for the male-specific melanization of these tergites in adults (Celniker et al. 1990; Jeong et al. 2006a). Unlike *D. melanogaster*, *D.*

pseudoobscura and *D. willistoni* do not show male specific melanization, but sub-elements of the *yellow* 5’ intergenic region from these species drive sexually dimorphic expression in the abdomen (e.g., will_C4). It is not possible to understand the full picture with the current data, but, if Abd-B is a real binder of *D.pseudoobscura* and *D. willistoni yellow* 5’ intergenic regions, perhaps the effects of Abd-B leading to male-specific *yellow* expression in tergites A5 and A6 are counteracted by neighboring transcription factor binding sites.

There is no previous knowledge about a direct relationship between *dsx* and *yellow*. In my study, Dsx was found to interact with two *D. willistoni yellow* enhancer sub-elements (will_C4 and will_C5), but also showed a pigmentation phenotype in *D. melanogaster* when knocked down with RNAi. Dsx has been shown to alter pigmentation through directly regulating another pigmentation gene *bab,* where it suppressed male-specific pigmentation in abdominal tergites A5 and A6 in females (Williams et al. 2008). Y1H now presents Dsx as a candidate binder of *yellow* 5’ intergenic region in *D. willistoni* and given the false negative rate of Y1H, this may be true for *D. pseuodoobscura* and *D. melanogaster* as well. Further *in vitro* binding and *in vivo* functional tests can shed light onto these possibilities.

Fru was found to interact with a *D. willistoni yellow* enhancer sub-element (will_C2), but affected pigmentation in *D. melanogaster*. This is, again, perhaps due to high false negative rate of Y1H and Fru is actually an interactor of *yellow* enhancers in *D. melanogaster* (and even *D. pseudoobscura*) as well. The effect of lack of Fru on pigmentation is particularly intriguing because this is the first time it was documented and it caused a bigger reduction in pigmentation in females than males. Previous studies showed that *fru* is genetically upstream of *yellow* in *D. melanogaster* males, albeit not related to pigmentation, but mating behavior. A male isoform of Fru (FRU^M) and Yellow were found to colocalize in the 3rd instar larval brain (Drapeau et al. 2003) and mutants that lack the DNA binding domain of FRU^M show reduced levels of of Yellow in male CNS only

(Drapeau et al. 2003). The same mutant line showed wild-type pigmentation in the body parts that harbor Yellow-induced pigmentation (e.g., abdomen, thorax, bristles, etc). The RNAi line used in my study appears to be targeting both male and female splice forms of Fru. With that in mind, it is perhaps not surprising to see a barely noticeable reduction of abdominal pigmentation in males upon knock-down of Fru (Drapeau 2006). The function of the female isoform of Fru is not known, but based on its sequence it appears to have a dimerization and a zinc-finger like DNA binding domain. Moreover, in females, *fru* mRNA is found at very low levels in the CNS and lack of *fru* does not appear to have any behavioral effects (Baker et al. 2001). Given that knock down of Fru in females leads to reduced abdominal pigmentation, perhaps the female isoform of this transcription factor is important, not for behavior but for the pigmentation pathway, possibly through affecting *yellow* expression.

Last, but not least, *vvl* was identified as a gene with a novel effect on pigmentation in *D. melanogaster* even though Y1H identified it to bind sub-element will_C2. Previous studies showed that lack of this gene prevents differentiation of longitudinal veins 2 and 4 which form on the ventral surface of the wing (Diaz-Benjumea et al. 1989). Overall, Vvl has been shown to be crucial for wing vein differentiation (de Celis et al. 1995), proper development of the tracheal system (Llimargas and Casanova 1997) and specific structures in the central nervous system such as the embryonic brain (Meier et al. 2006). It is expressed throughout development, with a peak during embryonic stages. It is known be expressed throughout the wing during pupa. It is downstream of genes such as *wingless* and *decapentaplegic* (dpp) (de Celis et al. 1995), which have been shown to specify (together with epidermal growth factor) the dorso-ventral pigmentation patterning of the abdominal segments in Drosophila by regulating downstream genes like *yellow* (Kopp et al. 1999; Wittkopp et al. 2003). Moreover, Vvl has been shown to be a direct binder of a 5' *shortvein* enhancer that drives *dpp* expression in the wing veins (Sotillos and de Celis 2006) reinforcing that it is involved in the same regulatory signaling pathways as *yellow*. In my study, Vvl was found to interact with a *D. willistoni yellow* enhancer sub-element (will_C2), but showed a pigmentation phenotype in *D. melanogaster*. This may again be due to the false negative rate of Y1H and perhaps Vvl binds to *D. melanogaster yellow* enhancers as well. Regardless, previous studies combined with my current data suggest roles for Vvl in abdominal pigmentation since it is downstream of the master regulators of this trait (Dpp and Wingless), is found to bind to *yellow* 5' intergenic harboring epidermal cell enhancer and upon knock down it leads to reduced abdominal pigmentation. Vvl is perhaps important for wing vein expression of *yellow* as well, since it is known to be expressed in the wing-veins throughout pupal stage. We do not yet know if sub-element will_C2 harbors wing vein enhancer activity. Further functional tests can show us how Vvl is involved in the regulation of *yellow* and the pigmentation pathway.

Enhancers with similar expression patterns can be responsive to similar or dissimilar sets of transcription factors

In understanding how enhancers change over evolutionary time, it is important to address whether enhancers with similar activities are bound by similar or different sets of transcription factors. If different, are there only few differences in the set or is the whole set different? As a result of Y1H, I defined a large set of transcription factors binding to different *yellow* enhancer sub-elements, which allowed me to make the aforementioned comparison. The preliminary comparison between the transcription factor binding profiles of *yellow* enhancer sub-elements conducted in my study was able to give one correlation. That is, I observed a similarity of transcription factor binding profile between half (five out of ten) of the enhancer sub-elements that drive male-specific expression pattern in abdominal tergites A5 and A6. This type of enhancer activity was fairly common among sub-elements tested (ten out of 21 showed it), and it is intriguing to find that it is a result of similar sets of transcription factors half of the time, but not all of the time (five out of ten had diverse transcription factor binding profiles).

My comparison remains preliminary, however, because the two datasets, activity and binding factors of *yellow* enhancer sub-elements, used in my study have aspects that are yet to be completed. Ideally, one needs to conduct a quantitative analysis of the complex expression patterns driven by enhancer sub-elements at multiple developmental stages to get a precise understanding of which activities are more similar to each other. It is also necessary to have a more comprehensive set of interactions between transcription factors and *yellow* enhancer sub-elements, because Y1H, with a high false negative rate, is likely to have not identified all true interactions. Without such a comprehensive set of transcription factors one cannot deduce how similar and different the transcription factor binding profiles of two enhancers are.

Materials and Methods

Constructing reporter genes

Previously identified 5' intergenic and and intronic regions of *yellow* from *D. melanogaster*, *D. pseudoobscura* and *D. willistoni* (Chapter-2) were dissected into approximately 1000 bp sub-elements using PCR (except for mel_A5, pse_B6, will_C7, which varied between 423 bp, 641 bp, 345 bp, respectively). Each sub-element overlapped with the flanking ones by approximately 100 bp (Sub-elements at the 5' and 3' ends of a 5' intergenic or intronic region overlap only with the element that is following or preceding them, respectively). PCR was conducted using a mix of Taq DNA polymearse and Phusion High-Fidelity DNA Polymerase (New England Biolabs) in order to prevent PCR-introduced mutations. Asc-1 restrictions enzyme site was introduced to the ends of each PCR product using primers with 5' Asc-1 tails. Subsequently, the PCR products for *yellow* enhancer sub-elements were subcloned into the sequencing vector pGEM-T, and sequenced using M13 Forward and M13 Reverse primers. Sequence-confirmed *yellow* enhancer sub-elements were then sub-cloned into a piggyBac-attB vector (as described in Chapter-2) using the Asc-1 unique site, which was followed by cloning nuclear Enhanced Green Fluorescent Protein (nEGFP) gene 3’ to each sub-element using the Fse-1 unique restriction site. The resulting construct was diagnostic digested and upon confirmation was prepared in high concentration using Zyppy Plasmid Maxi kit, reconfirmed with diagnostic digest and sent to Genetic Services, Inc, Cambridge, MA for injections into the attP-40 line of *D. melanogaster* (as described in Chapter-2)

Analysis of reporter gene expression patterns

For each line, transformant flies were crossed into w1118 line in order to ensure hemizygosity of the transgenes. Subsequently, pupa hemizygous for the transgenes were prepared 70-80 APF and imaged immediately using Leica SP5 confocal microscope as described previously (Chapter-2). All images were processed identically in Adobe Photoshop CS4.

Screening yellow *enhancer sub-elements for binding factors using yeast-one-hybrid (Y1H)*

Sub-elements of *yellow* tested for enhancer activity were cloned into Y1H-compatible pMW2 (“*HIS3*”) vector using Gateway cloning system (Invitrogen). In brief, *yellow* sub-elements were amplified with PCR using primers that had 5’ overhang attB sequences (25 bp) that were compatible with the attP site in the pDONR vector. A BP reaction was used to subclone the enhancer sub-element into the pDONR vector to create an Entry clone. The Entry clones were sequence confirmed which was followed by LR reactions to subclone the enhancer subelements into Y1H compatible pMW2 (“*HIS3*”) vector. These final constructs were mini-prepped and transformed (using lithium acetate (LiAc) - polyethylene glycol (PEG)) into the Y1H-aS2 yeast strain, where they were integrated into the mutant *his3-200* locus (Hens et al. 2011). Transformants were selected on SC -His, -Ura plates.

Four out of the 29 sub-elements were not assayed with Y1H due to lack of unique restriction enzyme site that was necessary to linearize the vector harboring the DNAbait::*HIS3* fusion construct prior to genome integration. The four sub-elements that were not assayed with Y1H correspond to *D. pseudoobscura* and *D. willistoni yellow* introns (pse_E2, pse_E3, will_F1 and will_F4).

In order to test whether the sub-elements of *yellow* enhancers integrated into the yeast genome activate expression of the HIS3 gene in the absence of any *D. melanogaster* transcription factors, i.e., under the influence of endogenous *S. cerevisiae* transcription factors, a self-activation test was conducted. For each of the 25 sub-elements that were integrated into the yeast genome, 8 transformants were picked and spotted onto plates containing varying concentrations of 3-Amino-1,2,4-triazole (3-AT) (0 mM to 100 mM) , a competitive inhibitor of the His3 enzyme. In order to be able to grow on increasing concentrations of 3-AT, transformants need to make higher amounts of the His3 enzyme than what is provided by basal level expression of *HIS3* gene. For all 25 sub-elements, among the 8 transformants spotted, one that was not able to grow on plates with 10, 20 or at the most 40 mM of 3-AT was selected for use in the rest of the experiments.

Subsequent yeast transformations to put each of the 647 the *D. melanogaster* transcription factors (prey) and 25 *yellow* enhancer sub-elements (bait) in the same cellular environment, yeast transformations were conducted in 384-well format using the LiAc-PEG method (Hens et al. 2011). The transformants in 384-well format were later on arrayed and converted to 1536-well format quadrupling each transformant strain (Figure-3) Subsequently the 1536-well format transformant plates were replica plated onto at least three increasing concentrations of 3-AT. On a 3-AT plate, a quadrant of transformants that grew above background levels indicate higher than basal level expression of the *HIS3* gene under the influence of the *D. melanogaster* transcription factor-Gal4AD fusion protein. This implies binding of the particular *D. melanogaster* transcription factor to the *yellow* enhancer sub-element being tested (Figure 3). All transformation plates were incubated at 30°C. Both 384-well and 1536-well format transformation plates were imaged after 3 days of growth using BioRad geldoc camera. The transformation plates containing 3-AT were imaged after 7 days as well as 10 days of growth at 30°C.

In order to conduct an unbiased analysis of all 3-AT plates for all of the 25 *yellow* enhancer sub-elements tested, the 3-AT plates were imaged and the images were analyzed using a MatLab program called TIDY (Transcription factor-DNA Interaction Detection in Yeast) (Hens et al. 2011). In a plate image, TIDY assesses the background level of colony growth and determines the quadrant of colonies that show growth above background levels. It also filters out the quadrants of colonies where there is non-uniform growth, i.e., only one or two out of the four colonies show growth above background levels. As a result it marks quadrants of colonies that show uniform and above background levels of growth as potential interaction between the particular bait being assayed and the transcription factor that correspond to that particular quadrant.

Using Cluster 3.0 to group yellow *enhancer sub-elements based on the similarity of their transcription factor binding profiles*

Cluster 3.0 (de Hoon et al. 2004) is a software designed to analyze microarray data by conducting various types of clustering. Even though it can analyze substantially more complex datasets, in this study Cluster was used to group the *yellow* enhancer sub-elements tested based on the similarity of their transcription factor binding profiles. In doing so the strength of interaction between a sub-element and transcription factor was ignored, and presence of interaction was designated by the number “1”, whereas the absence of interaction was designated by the number “0”. In the program, default settings were used and hierarchical clustering with centroid linkage was applied to the data set. The clustering results were visualized using Java Tree View (Alok Saldanha - http://jtreeview.sourceforge.net/)

Testing the in vivo effects of the identified transcription factors in D. melanogaster *using RNAi*

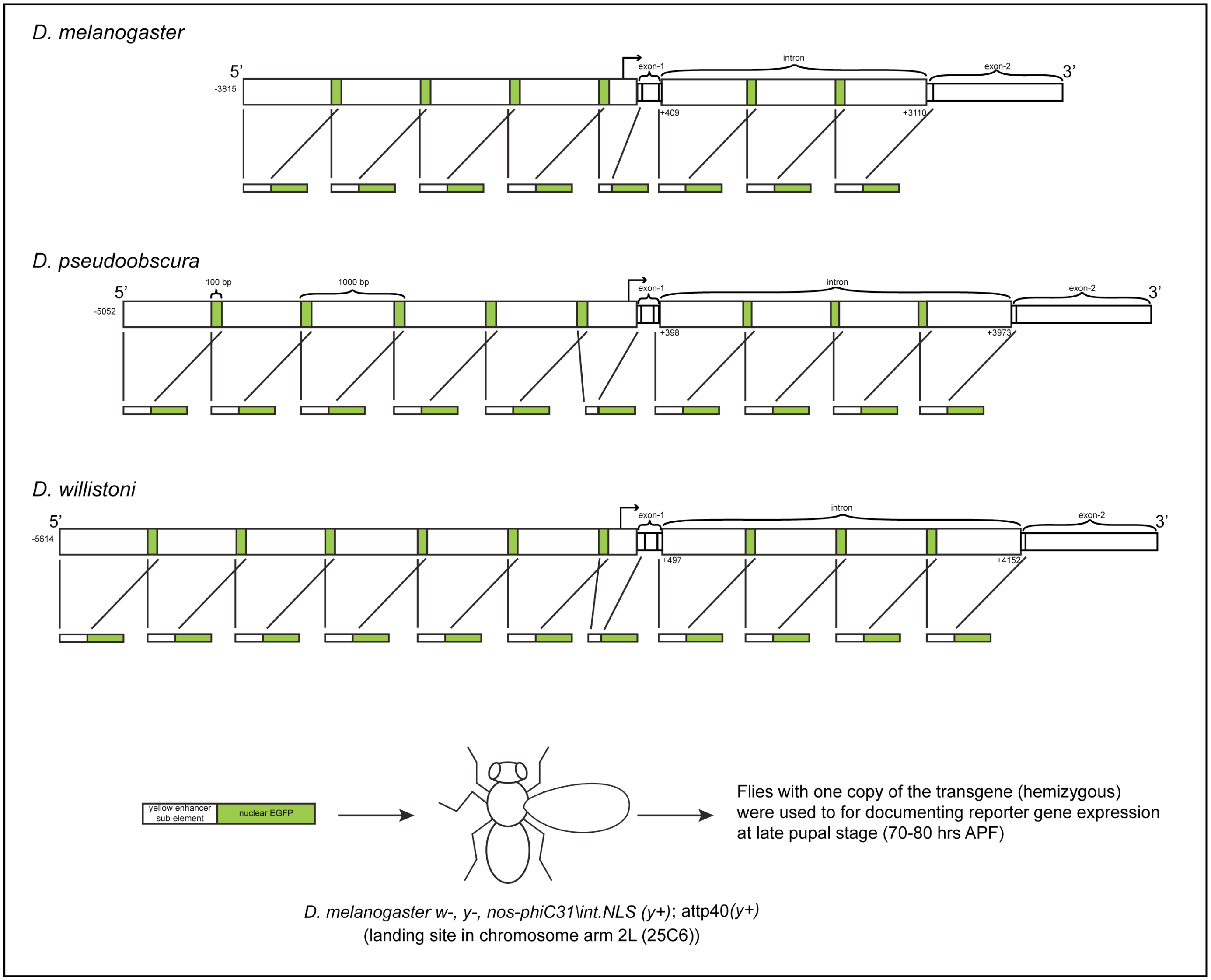
In order to test whether the transcription factors, which were found to bind various *yellow* enhancer sub-elements through Y1H, affect black pigment formation in adults, the UAS-GAL4 system was used to knock-down transcription factors, where when UAS-RNAi transgenes are crossed to a strain harboring a GAL4 driver, the RNAi is expressed in a pattern comparable to that of the GAL4 driver. Hence, RNAi lines for 40, out of the 204, transcription factors were obtained from Bloomington Drosophila Stock Center (Table 3-2 for the full list of RNAi lines used). All of the RNAi lines belonged to the TRIP (Transgenic RNAi Project, Harvard Medical School, NIH/NIGMS R01-GM084947) collection, where RNAi transgenes are integrated in the genome in a site directed manner (using the phiC31 mediated integration) rather than randomly, which prevents position effects on RNAi expression. Moreover, the vectors harboring the RNAi transgenes have multiple UAS sites 5’ to the RNAi sequence, introns within RNAi transgenes, SV40 polyadenylation signal at the 3’ end, and insulators flanking the RNAi element, all of which helps make the expression of the RNAi transgene consistent and strong when activated. Hence upon activation, RNAi lines constructed using these vectors are expected to result in more robust knock-downs than randomly inserted RNAi transgenes (Ni et al. 2008; Ni et al. 2009). This allowed choosing sites with low basal activity and also made this basal activity consistent among RNAi lines their expression consistent. In this study, all RNAi transgenes used were integrated in the attP2 site on the third chromosome.

The crosses between specific RNAi lines and the line harboring the *pnr*-Gal4 driver were carried out as follows: virgin females from RNAi lines with the genotype *y, sc, v*; attP2 (y^1, sc^1, v^1; P{y[+t7.7]=CaryP}attP2), P{*UAS-RNAi* y^+ v^+} were crossed to males with the genotype y^1, w^{1118}; P{GawB}pnrMD^{237}/TM3, P {UAS-y.C}MC2, Ser^1. In the progeny females that have the *y, sc, v* /y^1, w^{1118} ; attP2 (y^1, sc^1, v^1; P{y[+t7.7]=CaryP}attP2), P{*UAS-RNAi* y^+ v^+} / P{GawB} pnrMD^{237} and males that have genotype *y, sc, v* ; attP2 (*y^1*, *sc^1*, *v^1*; P{y[+t7.7] =CaryP}attP2), P{*UAS-RNAi y^+ v^+*} / P{GawB}pnrMD^{237} were selected based on the red eye color (rescue of the vermillion mutation), which indicates presence of the RNAi transgene, and lack of humeral phenotype (travels with TM3), which indicates the presence of the *pnr*-Gal4 transgene. These flies were the "test" flies since in the presence of the Gal4 driver they should express the RNAi transgene and the knock-down of the transcription factor is expected to occur. On the other hand "control" flies (*y, sc, v* /*y^1, w^{1118}* ; attP2 (*y^1*, *sc^1*, *v^1*; P{y[+t7.7]=CaryP}attP2), P{*UAS-RNAi y^+ v^+*} /TM3 in females and *y, sc, v* ; attP2 (*y^1*, *sc^1*, *v^1*; P{y[+t7.7] =CaryP}attP2), P{*UAS-RNAi y^+ v^+*} / TM3 in males) had red eyes (presence of RNAi transgene), but also showed the humeral phenotype and, hence, were not expected to express the RNAi transgene or a related phenotype. All flies were raised at 20°C. Progeny from the aformentioned crosses were collected everyday, sorted based on sex as well as eye color and humeral phenotypes, aged 3-5 days, and put in 1:10 Glycerol:Ethanol mix. After being kept in Glycerol:Ethanol mix at least for 3 days, abdominal cuticle of flies from all four lines where altered pigmentation was observed as well as of flies from two other lines that did not appear to show a pigmentation phenotype were dissected, mounted in polyvinyl alcohol mounting medium (Sigma-Aldrich), baked at 65°C overnight and imaged using Schott Leica mz-6 microscope camera and "Scion Visicapture" version 1.2 software.

Acknowledgements

I would like to thank Mackenzie Dome for help with cloning *yellow* enhancer sub-elements, fly crosses, pupal dissections, cuticle dissections and imaging, Elizabeth Walker for help with DNA preps, Bart Deplancke, Korneel Hens, Jean-Daniel Feuz, and Antonina Iagovitina for help with yeast-one-hybrid data collection and analysis.

Figure 3-1



Flies with one copy of the transgene (hemizygous) were used to for documenting reporter gene expression at late pupal stage (70-80 hrs APF)

D. melanogaster w-, y-, nos-phiC31\int.NLS (y+); attp40*(y+)*
(landing site in chromosome arm 2L (25C6))

Figure 3-1. Sub-elements of *yellow* 5' intergenic and intronic regions from three Drosophila species were tested for enhancer activity.
Schematics of *yellow* 5' intergenic and intronic regions from *D. melanogaster, D. pseudoobscura* and *D. willistoni*. Each 5' intergenic and intronic region was sub-divided into DNA elements ~1000 base pairs (bp) long, that overlapped with the flanking sub-element by ~100 bp. The resulting sub-elements were used to drive reporter gene (nuclear enhanced green fluorescent protein gene) expression in the attP40 line of *D. melanogaster*.

Figure 3-2

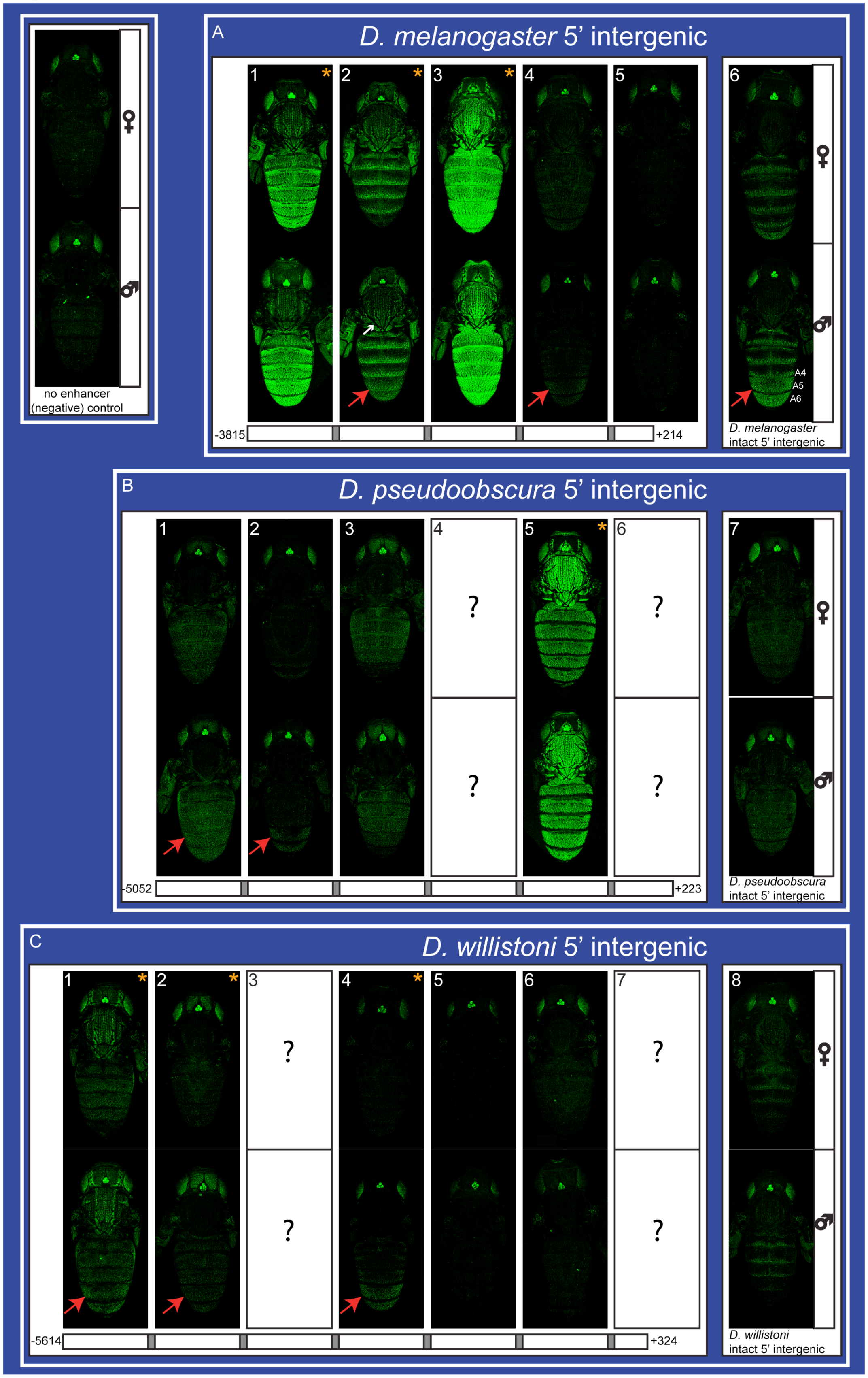

♀
♂
no enhancer
(negative) control
A
D. melanogaster 5’ intergenic
1 *
2 *
3 *
4
5
6
-3815
+214
A4
A5
A6
D. melanogaster
intact 5’ intergenic
B
D. pseudoobscura 5’ intergenic
1
2
3
4
?
?
5 *
6
?
?
7
-5052
+223
D. pseudoobscura
intact 5’ intergenic
C
D. willistoni 5’ intergenic
1 *
2 *
3
?
?
4 *
5
6
7
?
?
8
-5614
+324
D. willistoni
intact 5’ intergenic

Figure 3-2

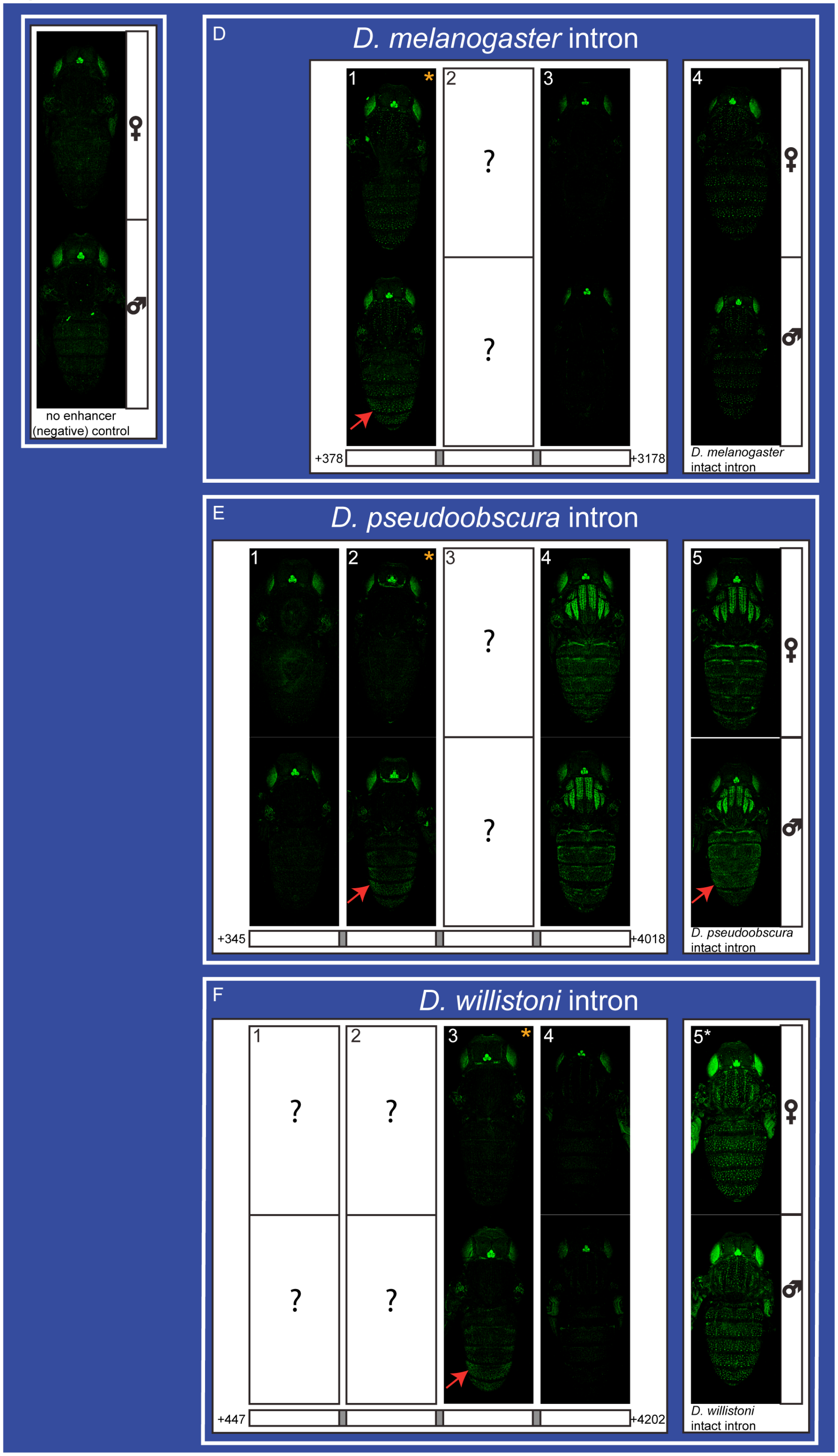

no enhancer
(negative) control
♀
♂
D
D. melanogaster intron
1 *
2
3
4
?
?
+378
+3178
D. melanogaster
intact intron
E
D. pseudoobscura intron
1
2 *
3
4
5
?
?
+345
+4018
D. pseudoobscura
intact intron
F
D. willistoni intron
1
2
3 *
4
5*
?
?
?
?
+447
+4202
D. willistoni
intact intron

Figure 3-2. *yellow* enhancer sub-elements drive diverse expression patterns some of which are “cryptic”.

Confocal images of pupa harboring one copy (hemizygous) of a transgene where a 5’ intergenic or intronic *yellow* enhancer sub-element drives expression of the reporter gene (nEGFP in green). GFP expression in the eye and ocelli is driven by the transformation marker (3xp3-EGFP) and not a *yellow* enhancer sub-element. The negative control line harbors a transgene where nEGFP is not driven by a putative enhancer (basal level expression). Red arrows point to the male-specific expression in tergites A5 and A6 (sometimes A4) pattern in ten transgenic lines. White arrow points to the expression in the scutellum in mel_A2. Abdominal tergites A4, A5 and A6 are indicated in panel 2A6. Orange asterisks (*) indicate the ten enhancer sub-elements that drove expression in spatial patterns that were not part of the pattern driven by the full region they were isolated from (cryptic enhancer activity). The white asterix (*) indicates he transgenic line harboring the full *D. willistoni yellow* intron was lost and hence could not be imaged in hemizygous from. A previous image of a homozygous fly from the same line is used in place. Question marks indicate the sub-elements for which a trangenic line has not been acquired yet.

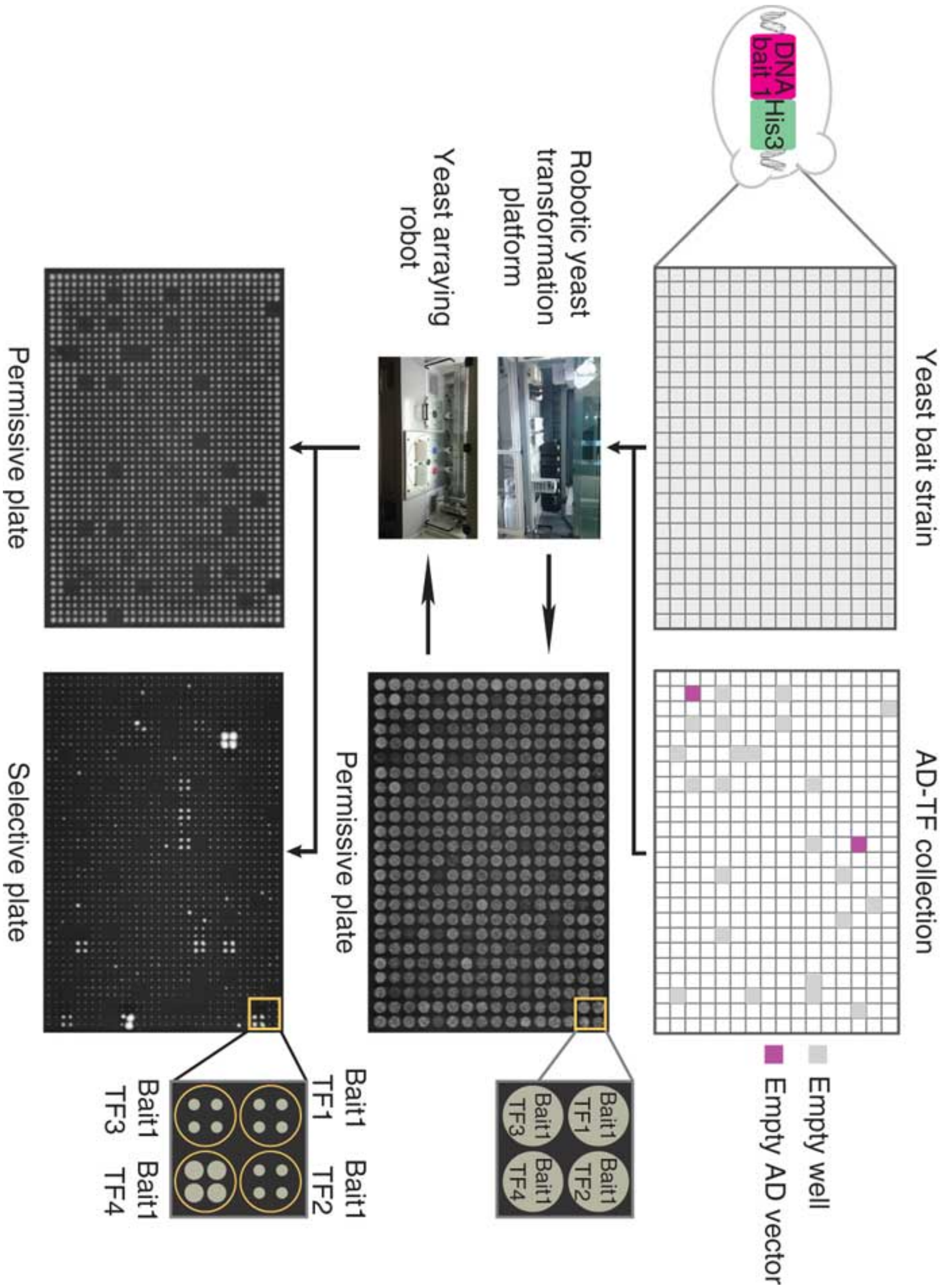

DNA bait 1
His3
Yeast bait strain
AD-TF collection
Empty well
Empty AD vector
Robotic yeast transformation platform
Yeast arraying robot
Permissive plate
Permissive plate
Selective plate
Bait1 TF1
Bait1 TF2
Bait1 TF3
Bait1 TF4

Figure 3-3

Figure 3-3. (Courtesy of collaborator Bart Deplancke and his lab) The Drosophila high-throughput Y1H platform.

A yeast DNA bait strain is distributed over a 384-well plate. Each well of this plate is then transformed with a different Gal4-Activation Domain(AD)-transcription factor clone from the *Drosophila* Y1H AD-transcription factor library by a robotic yeast transformation platform which spots the 384 individually transformed yeast strains on a permissive agar plate. A colony- pinning robot subsequently transfers the yeast colonies onto a permissive and a selective plate, quadruplicating each colony in a square pattern in the process. Transcription factor- DNA bait interactions are identified based on growth on a selective, 3-AT-containing yeast plate

Figure 3-4

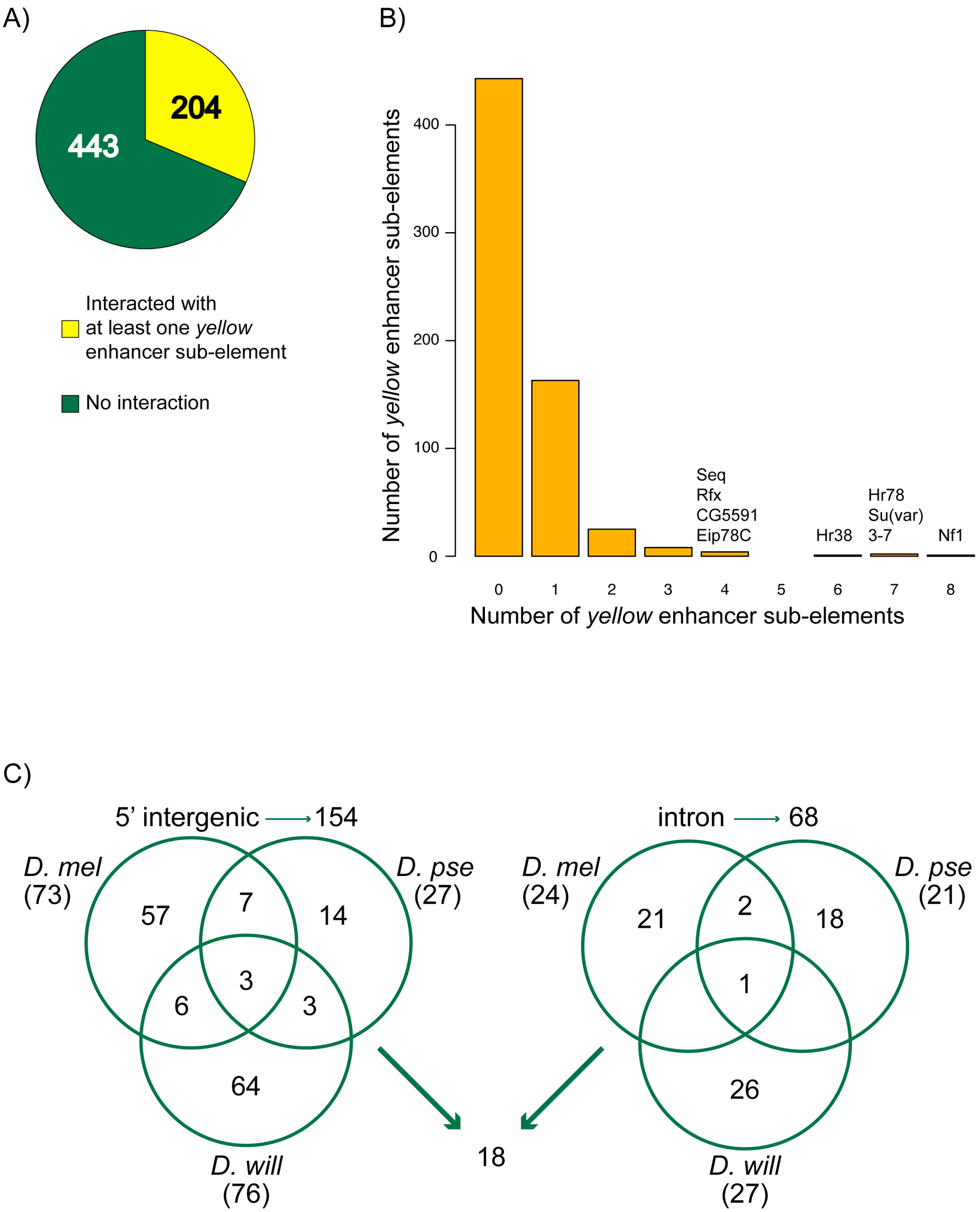

A)
204
443
Interacted with at least one *yellow* enhancer sub-element
No interaction
B)
Number of *yellow* enhancer sub-elements
400
300
200
100
0
Seq
Rfx
CG5591
Eip78C
Hr38
Hr78
Su(var)
3-7
Nf1
0
1
2
3
4
5
6
7
8
Number of *yellow* enhancer sub-elements
C)
5' intergenic → 154
D. mel
(73)
D. pse
(27)
57
7
14
3
6
3
64
D. will
(76)
18
intron → 68
D. mel
(24)
D. pse
(21)
21
2
18
1
26
D. will
(27)

Figure 3-4. Distributions of the transcription factors identified by Y1H as candidate binders of *yellow* 5' intergenic and intronic enhancers

A) Out of the 647 D. melanogaster tested with Y1H, only 204 were shown to bind to at least one *yellow* enhancer sub-element from any of the three Drosophila species studied.

B) Histogram showing the number of transcription factors that were found to bind 0, 1, 2, 3, 4, 5, 6, 7, 8 *yellow* enhancer sub-elements. Names of the transcription factors that were found to bind 4 or more *yellow* enhancer sub-elements are indicated above their corresponding column.

C) Venn diagrams showing the number of shared and unique binders between physically homologous regions in different species. On the left, a total of 154 transcription shown to bind 5' intergenic regions of *yellow* from *D. mel* (*D. melanogaster*), *D. pse* (*D. pseudoobscura*), and *D. will* (*D. willistoni*) were distributed among the three species. On the right a total of 68 transcription shown to bind introns *yellow* from *D. mel*, *D. pse*, and *D. will* were distributed among the three species. Overall there are 18 transcription factors shared between 5' intergenic and intronic regions.

Figure 3-5

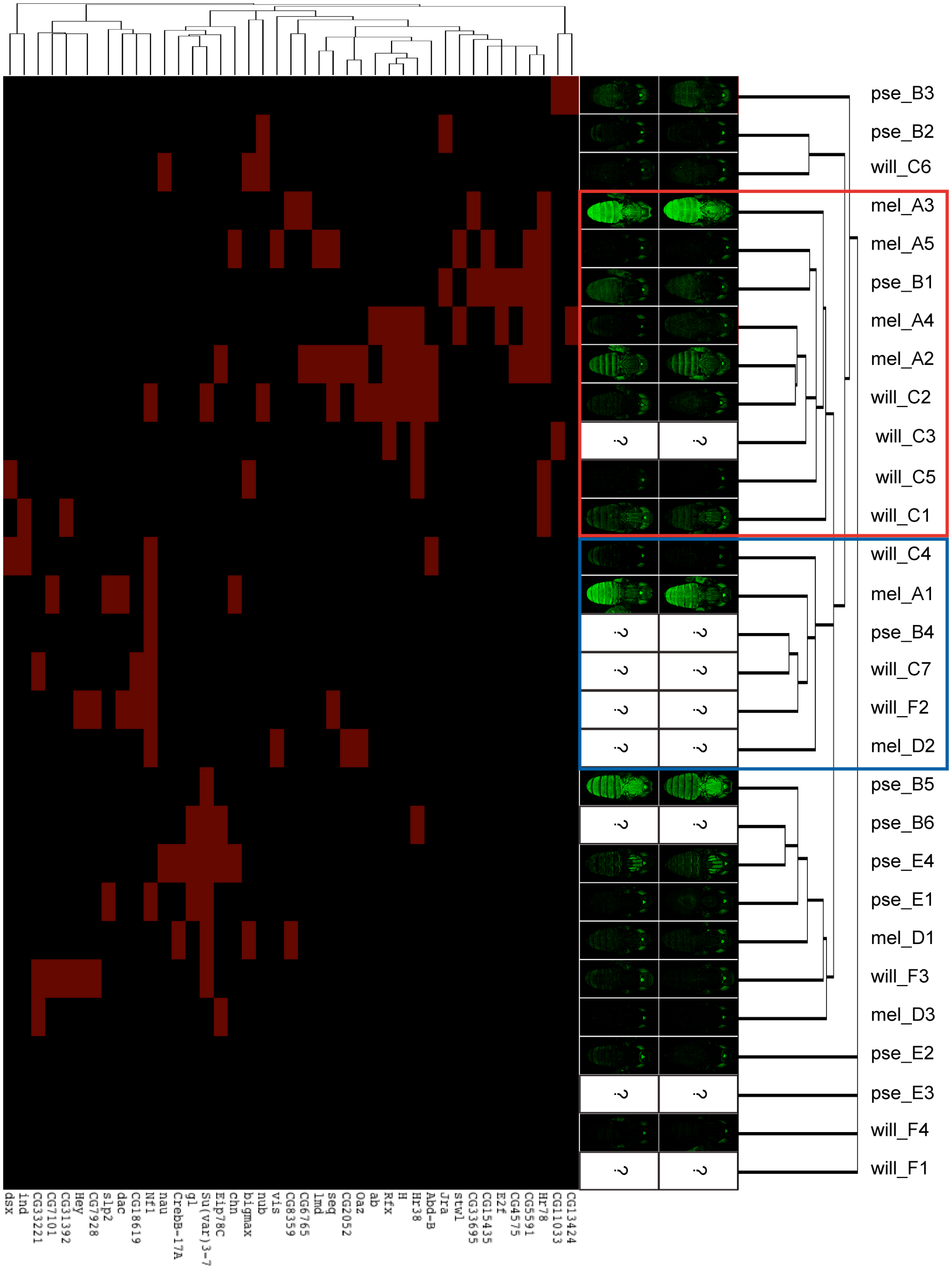

pse_B3
pse_B2
will_C6
mel_A3
mel_A5
pse_B1
mel_A4
mel_A2
will_C2
will_C3
will_C5
will_C1
will_C4
mel_A1
pse_B4
will_C7
will_F2
mel_D2
pse_B5
pse_B6
pse_E4
pse_E1
mel_D1
will_F3
mel_D3
pse_E2
pse_E3
will_F4
will_F1
?
dsx
ind
CG33221
CG7101
CG31392
Hey
CG7928
slp2
dac
CG18619
Nf1
nau
CrebB-17A
gl
Su(var)3-7
Eip78C
chn
bigmax
nub
vis
CG8359
CG6765
lmd
seq
CG2052
Oaz
ab
Rfx
H
Hr38
Abd-B
Jra
stwl
CG33695
CG15435
E2f
CG4575
CG5591
Hr78
CG11033
CG13424

Figure 3-5. There is not a strong correlation between the transcription factor binding profiles and the activities of *yellow* enhancer sub-elements. Using the program Cluster 3.0, *yellow* enhancer sub-elements were grouped based on the similarity of their transcription factor binding profiles (of only the transcription factors (43 of them) that were shown to interact with more than one sub-element) (top tree). Transcription factors were grouped based on the similarity between the groups of *yellow* enhancer sub-elements they bind to (side tree). Red rectangle is around the group of *yellow* enhancer sub-elements, where five (out of the eight in the group) show sexually dimorphic enhancer activity in the abdomen. Blue rectangle is around a group of yellow enhancer sub-elements that appear to be bound by the transcription factor Nf1. Names of the transcription factors are listed on the right side of the cluster figure. *yellow* enhancer sub-elements were referred to as following from Figure 2, e.g., *D. melanogaster* 5' intergenic sub-elements are mel_A1, 2, 3, 4, 5. Questions marks indicate the sub-elements whose enhancer activity has not been identified yet.

Figure 3-6

Fly harboring *pannier*-Gal4, UAS-GFP

Figure 3-6C

Figure 3-6. Change in abdominal pigmentation upon knock down of six transcription factors

Control flies lack a Gal4 driver activating of UAS-RNAi transgene against the particular transcription factor. Knock-down flies harbor *pannier*-Gal4 driver that drives UAS-RNAi expression in the dorsal midline, which is where change in pigmentation is expected it to be observed.

A) *pannier* expression pattern in the dorsal midline of a fly carrying *pannier*-Gal4 and UAS-GFP transgenes.

B) Abdominal cuticle dissections of flies harboring UAS-RNAi constructs for the four transcriptions that appear to affect pigmentation. Abd-B (Abdominal-B), Dsx (Doublesex), Fru (Fruitless), Vvl (Ventral veins lacking). Red arrows point to regions showing altered pigmentation.

C) Abdominal cuticle dissections of flies harboring UAS-RNAi constructs for two of the many transcription factors that did not appear to affect pigmentation when knocked down with a single UAS-RNAi line. Hr38 (Hormone receptor like in 38), Hr78 (Hormone receptor like in 78) were shown to bind 6 and seven *yellow* enhancer sub-elements, respectively.

Table 3-1 *D. melanogaster* transcription factors that were found to activate *yellow* enhancer sub-elements in a yeast-one-hybrid assay

D. melanogaster trancription factors that activated at least one 1-kb enhancer fragment	*D. melanogaster yellow* 5' intergenic region					*D. pseudoobscura yellow* 5' intergenic region						*D. willistoni yellow* 5' intergenic region							*D. melanogaster yellow* intron			*D. pseudoobscura yellow* intron				*D. willistoni yellow* intron			
	mel_A1	mel_A2	mel_A3	mel_A4	mel_A5	pse_B1	pse_B2	pse_B3	pse_B4	pse_B5	pse_B6	will_C1	will_C2	will_C3	will_C4	will_C5	will_C6	will_C7	mel_D1	mel_D2	mel_D3	pse_E1	pse_E2 (no Y1H data)	pse_E3 (no Y1H data)	pse_E4	will_F1 (no Y1H data)	will_F2	will_F3	will_F4 (no Y1H data)
Hr78		■	■	■	■	■						■				▨													
Hr38		■		■							■		■	▨		▨													
Nf1	■								■				■		■			■		▨		▨					■		
CG5591		■		■	▨	▨							■		■			■											
Eip78C		■									▨										▨				■				
CG4575		■				■																							
CG13424				■				■																					
E2f				▨		▨																							
CG33695			▨			■																							
CG15435					■	■																							
Abd-B		▨											■		▨														
seq		▨			■								▨														■		
H		▨		▨									▨																
Rfx		▨		▨									■	▨															
Oaz		▨											▨							■									
ab				■									▨																
nub							■						▨				▨												
CG11033								■						▨															
Su(var)3-7										■	■		■						■			■			■			■	
Imd		▨			■																								
CG6765		▨	▨																										
stwl				▨	▨																								
chn	■				▨																				■				
Jra						■	■																						
ind												▨			■														
bigmax																▨	▨		■										
CrebB-17A																			▨						▨				
gl											■											▨			▨				
CG7928																											■	■	
Hey																											■	■	
CG2052		▨																		▨									
CG8359			■																▨										
vis					▨															▨									
slp2	■																					■							
CG7101	■																											■	
dac	■																										■		
CG31392												■																■	
CG33221																		■			■							■	
CG18619																		■									■		
dsx															▨	■													
nau																	■								▨				
CG1233	■																												
NFAT	■																												
CG4936	▨																												
YL-1	▨																												
CG15710	▨																												
foxo	▨																												
sima	▨																												
CG16779		■																											
Psc		■																											
CG11966		■																											
ftz-f1		■																											
mld		■																											
CG17829		▨																											
scrt		▨																											
CG18599		▨																											

Table 3-1 *D. melanogaster* transcription factors that were found to activate *yellow* enhancer sub-elements in a yeast-one-hybrid assay

D. melanogaster trancription factors that activated at least one 1-kb enhancer fragment	*D. melanogaster yellow* 5' intergenic region					*D. pseudoobscura yellow* 5' intergenic region						*D. willistoni yellow* 5' intergenic region							*D. melanogaster yellow* intron			*D. pseudoobscura yellow* intron				*D. willistoni yellow* intron			
	mel_A1	mel_A2	mel_A3	mel_A4	mel_A5	pse_B1	pse_B2	pse_B3	pse_B4	pse_B5	pse_B6	will_C1	will_C2	will_C3	will_C4	will_C5	will_C6	will_C7	mel_D1	mel_D2	mel_D3	pse_E1	pse_E2 (no Y1H data)	pse_E3 (no Y1H data)	pse_E4	will_F1 (no Y1H data)	will_F2	will_F3	will_F4 (no Y1H data)
dl			■																										
mamo			■																										
CG1647			■																										
dimm			▨																										
kn			▨																										
org-1			▨																										
CG33396				■																									
ush				■																									
Hmgz				■																									
brm				■																									
Hmx				■																									
CG9932				■																									
CG9727				▨																									
CG7987				▨																									
CG30020				▨																									
CG17361				▨																									
psq				▨																									
HLH106				▨																									
CG15011				▨																									
D19A				▨																									
CG14667				▨																									
Med				▨																									
grau				▨																									
dei				▨																									
disco-r				▨																									
CG31510				▨																									
CG6791					■																								
CenP-C					■																								
Chrac-14					■																								
tef					■																								
CG11617					▨																								
Her					▨																								
Oli						■																							
Snoo						■																							
BEAF-32						▨																							
CG8944						▨																							
CG6686						▨																							
ato						▨																							
HLH4C							■																						
CG9650							■																						
Kr-h1									■																				
dve										■																			
bowl										■																			
Lmpt											■																		
Ro												■																	
Gsc												■																	
CG17806												■																	
HGTX												■																	
Scr												■																	
CG17287												■																	
net												■																	
CG2790												■																	
Su(z)12												■																	
CG17836												■																	
CG11085												▨																	
retn												▨																	

Table 3-1 *D. melanogaster* transcription factors that were found to activate *yellow* enhancer sub-elements in a yeast-one-hybrid assay

D. melanogaster trancription factors that activated at least one 1-kb enhancer fragment	*D. melanogaster yellow* 5' intergenic region					*D. pseudoobscura yellow* 5' intergenic region						*D. willistoni yellow* 5' intergenic region							*D. melanogaster yellow* intron			*D. pseudoobscura yellow* intron				*D. willistoni yellow* intron			
	mel_A1	mel_A2	mel_A3	mel_A4	mel_A5	pse_B1	pse_B2	pse_B3	pse_B4	pse_B5	pse_B6	will_C1	will_C2	will_C3	will_C4	will_C5	will_C6	will_C7	mel_D1	mel_D2	mel_D3	pse_E1	pse_E2 (no Y1H data)	pse_E3 (no Y1H data)	pse_E4	will_F1 (no Y1H data)	will_F2	will_F3	will_F4 (no Y1H data)
CG9797												▨																	
AP-2												▨																	
HLHm&ggr												▨																	
cas													■																
vvl													▨																
E(bx)													▨																
hang													▨																
CG31782													▨																
CG2120													▨																
lbl													▨																
fru													▨																
p53														■															
ovo														▨															
CG32772														▨															
sd														▨															
rgr														▨															
pros														▨															
jing														▨															
CG30431															■														
C15															■														
Lim3															■														
pdm2															■														
otp															▨														
h																▨													
CG13287																▨													
yem																▨													
Dref																▨													
CG1845																	■												
CG4854																	■												
CG2678																	■												
pnt																	■												
CG11984																	■												
CG10431																	■												
Tip60																	■												
esc																	▨												
CG17803																	▨												
CG8765																	▨												
l(1)sc																	▨												
en																		■											
Dsp1																		■											
CG18619																		■											
Hp1c																		▨											
CG15073																		▨											
H2.0																		▨											
Hr4																		▨											
Doc2																			■										
caup																			■										
cyc																			■										
crm																			■										
CG3815																			▨										
cg1529																				■									
cg9437																				■									
gt																				■									
CG11676																				▨									
CG17195																				▨									
CG10462																				▨									

Table 3-1 *D. melanogaster* transcription factors that were found to activate *yellow* enhancer sub-elements in a yeast-one-hybrid assay

D. melanogaster trancription factors that activated at least one 1-kb enhancer fragment	*D. melanogaster yellow* 5' intergenic region					*D. pseudoobscura yellow* 5' intergenic region						*D. willistoni yellow* 5' intergenic region							*D. melanogaster yellow* intron			*D. pseudoobscura yellow* intron				*D. willistoni yellow* intron			
	mel_A1	mel_A2	mel_A3	mel_A4	mel_A5	pse_B1	pse_B2	pse_B3	pse_B4	pse_B5	pse_B6	will_C1	will_C2	will_C3	will_C4	will_C5	will_C6	will_C7	mel_D1	mel_D2	mel_D3	pse_E1	pse_E2 (no Y1H data)	pse_E3 (no Y1H data)	pse_E4	will_F1 (no Y1H data)	will_F2	will_F3	will_F4 (no Y1H data)
Blimp-1																				▨									
achi																				▨									
hb																					▨								
D1																					▨								
CG16801																						■							
eve																						■							
CG4282																						▨							
dmrt99B																						▨							
nej																									■				
Hr46																									▨				
NC2&agr																									▨				
Ets65A																									▨				
knri																									▨				
CG17359																									▨				
cro1																									▨				
bun																									▨				
pho																									▨				
CG11247																									▨				
SoxN																											■		
CG12744																											■		
toe																												■	
CG31666																												■	
lab																												■	
Aef1																												■	
CG2202																												■	
HLH54F																												■	
Dr																												■	
d4																												■	
CG9571																												■	
Met																												■	
CG18446																												■	
CG12054																												▨	
CG16863																												▨	
run																												▨	
CG4496																												▨	
Max																												▨	

Table 3-1. *D. melanogaster* transcription factors that were found to activate *yellow* enhancer sub-elements in a yeast-one-hybrid assay
List of *Drosophila melanogaster* transcription factors with the matching *yellow* enhancer sub-elements (from the 5' intergenic and intronic regions of *D. melanogaster*, *D. pseudoobscura* and *D. willistoni*) that they were found to activate in a yeast-one-hybrid assay. Blue shaded boxes indicate interaction between a particular transcription factor and *yellow* enhancer sub-element. Light shaded blue indicate interaction found in only one (out of three) test plates, dark shaded blue boxes indicate interaction found in more than one test plate. Transcription factors in pink font were among the ones that were frequently found to activate different enhancer elements tested by the Deplancke Lab, hence, they may be showing unspecific activation (personal communication). Transcription factors shaded in yellow were knocked-down using RNA interference in *D. melanogaster*.

	Table 3-2 Gene Ontology characteristics of transcription factors identified by yeast-one-hybrid to bind to *yellow* enhancers										
	Transcription factor that putatively binds to *yellow* enhancers	molecular function	biological process it is involved in	mutant phenotypes annotated with	expression	mel 5'	mel intron	pse 5'	pse intron	will 5'	will intron
1	Hr78	(Hormone-receptor-like in 78 --> ligand dep nuclear receptor activity)	autophagic cell death; liquid clearance, open tracheal system; open tracheal system development; salivary gland cell autophagic cell death	embryonic/larval trachea; embryonic/larval tracheal system	moderately high expression throughout development	4/5	NB	1/6	NB	2/7	NB
2	Hr38	(Hormone-receptor-like in 38 --> ligand dep nuclear receptor activity)	cuticle development; phagocytosis, engulfment.	alleles are annotated with: leg; adult epidermis; adult cuticle; pupal epidermis; joint; pupal cuticle	expression peaks at 2-day pupae to/through adult stage	2/5	NB	1/6	NB	3/7	NB
3	Nf1	Neurofibromin1 --> Ras GTPase activator activity	biological regulation; response to stimulus; learning or memory; response to stress; system process; intracellular signal transduction; regulation of multicellular organismal process; cognition; cellular component organization or biogenesis; associative learning; determination of adult lifespan; locomotory behavior; short-term memory	pupa; wing disc; surface associated glial cell; embryonic heart; eye photoreceptor cell; wing	moderate expresion throughout	1/5	1/3	1/6	1/4	3/7	1/2
4	CG5591	predicted: DNA binding; zinc ion binding	phagocytosis, engulfment	no data	high in embryo, moderate thorughout pupae and adult	x3	NB	x1	NB	NB	NB
5	Eip78C	Ecdysone-induced protein 78C --> ligand-dependent nuclear receptor activity	instar larval or pupal development	eye; ommatidium; dorsal appendage	sharply peaks to "high" expression on 2-day-old pupae	x1	x1	x1	x1	NB	NB
6	CG4575	predicted: sequence-specific DNA binding transcription factor activity; protein homodimerization activity	predicted: regulation of transcription, DNA-dependent	no data	sharply peaks to "low" expression from extremely low on 2-day-old pupae (highest in 2-4hr embryo and 2-d old pupae	x1	NB	x1	NB	NB	NB
7	CG13424	lateral muscles scarcer (lms) --> predicted: sequence-specific DNA binding transcription factor	larval somatic muscle development; flight	wing	moderate during embryo, steady "very low" throughout adult	x1	NB	x1	NB	NB	NB
8	E2f	E2F transcription factor --> protein binding; DNA binding	positive regulation of gene expression; cellular process; dendrite morphogenesis; neuron development; positive regulation of nurse cell apoptosis; muscle tissue development; antimicrobial humoral response; regulation of cell cycle; DNA endoreduplication	organ system subdivision; organ system; nervous system; peripheral nervous system; external compound sense organ; primordium; female germline cyst; egg; adult segment; imaginal precursor	highest in embryonic stages, moderate-to-moderate-high throughout the rest	x1	NB	x1	NB	NB	NB
9	CG33695	predicted: DNA binding	no data	no data	highest(mod high) in early embryo and adult female, moderate for the rest of the stages and adult males	x1	NB	x1	NB	NB	NB
10	CG15435	predicted: zinc ion binding	no data	no data	highest in early embryo and adult female, moderate for the rest of the stages and adult males	x1	NB	x1	NB	NB	NB
11	Abd-B	Abdominal B --> sequence-specific DNA binding transcription factor activity	anatomical structure development; organ development; biological regulation; multicellular organismal reproductive process; multicellular organismal process; open tracheal system development; development of primary male sexual characteristics; regulation of multicellular organismal development; genital disc development; cellular process involved in reproduction; segment specification; gonadal mesoderm development; regulation of developmental process	organ system subdivision; adult segment; organ system; abdominal ventral denticle belt; spiracle; late extended germ band embryo; embryonic segment; abdominal segment 7; embryonic abdomen; thoracic segment	highest(mod) 4-24 hr embryo, low for most of the rest, very low for 0-2 hr embryo, L3 1-2, pupa 4d, adult fem	x1	NB	NB	NB	x2	NB
12	seq	sequoia --> predicted: zinc ion binding; nucleic acid binding	negative regulation of axon extension; axon guidance; sensory organ development; dendrite morphogenesis; axonogenesis	no data	very high in 6-8-hr old embryo, moderate during larvae, pupae and adult	x2	NB	NB	NB	x1	

Table 3-2 Gene Ontology characteristics of transcription factors identified by yeast-one-hybrid to bind to *yellow* enhancers

	Transcription factor that putatively binds to *yellow* enhancers	molecular function	biological process it is involved in	mutant phenotypes annotated with	expression	mel 5'	mel intron	pse 5'	pse intron	will 5'	will intron
13	H	Hairless --> transcription corepressor activity	negative regulation of Notch signaling pathway; sensory organ boundary specification; regulation of cell death; sensory organ precursor cell fate determination; imaginal disc-derived wing margin morphogenesis; imaginal disc-derived wing vein morphogenesis; wing disc dorsal/ventral pattern formation	organ system subdivision; adult segment; peripheral nervous system; nervous system; adult mesothoracic segment; external compound sense organ; eo support cell; primordium; imaginal precursor; eo sensory structure; dorsal thoracic disc; embryonic hindgut; adult cuticle; external sensory organ precursor cell; adult external prothorax; late extended germ band embryo; postalar bristle	highest(mod high) until 10hr-embryo and adult females, low to moderate the rest of the stages and low in adult males	x2	NB	NB	NB	x1	NB
14	Rfx	Rfx --> predicted: DNA binding	sensory perception of sound; nervous system development	antenna; macrochaeta; abdominal lateral pentascolopidial chordotonal organ lch5	moderate at all stages except for late ebryo and all larvae where it is low	x2	NB	NB	NB	x2	NB
15	Oaz	O/E-associated zinc finger protein--> predicted: zinc ion binding; nucleic acid binding	spiracle morphogenesis, open tracheal system	filzkorper; embryonic/larval tracheal system; wing	highest(low) during embryo, larva, whiteprepupa, "very low" for the rest	x1	NB	NB	NB	x1	NB
16	ab	abrupt --> predicted: sequence-specific DNA binding transcription factor activity	border follicle cell migration; dendrite morphogenesis; neuron development; muscle organ development	hypodermal muscle of larval abdomen; organ system; adult segment; organ system subdivision; peripheral nervous system; larval abdominal segment; abdominal dorsal oblique muscle 3; hypodermal muscle of larval abdominal 2; sensory cluster; abdominal 6 oblique muscle	highest(high) during 6-16-hr embryo, moderately high during white prepupae, moderate during pupae and adult male	x1	NB	NB	NB	x1	NB
17	nub	nubbin --> sequence-specific DNA binding transcription factor activity	wing disc development	organ system subdivision; adult segment; adult antennal lobe; peripheral nervous system; organ system; embryonic/larval neuron; antennal olfactory receptor neuron; thoracic segment; integumentary specialisation; adult uniglomerular antennal lobe projection neuron adPN	high to modearte during embryo, moderate during white prepupa, low during pupae and adult	NB	NB	x1	NB	x2	NB
18	CG11033	Lysine (K)-specific demethylase 2 --> histone demethylase activity (H3-trimethyl-K4 specific)	histone H3-K4 demethylation, trimethyl-H3-K4-specific	nucleolus	highest (mod high) during 4-16-hr embryo and white prepupa, moderate for the rest	NB	NB	x1	NB	x1	NB
19	Su(var)3-7	Suppressor of variegation 3-7 --> predicted: DNA binding; zinc ion binding	dosage compensation, by inactivation of X chromosome	macrochaeta; humeral bristle; sex comb; wing; ommatidium	highest (high)during early embryo, mod high the rest of pupa, moderate during larva and pupa, mod high in adult fem	NB	x1	x2	X2	x2	X1
20	lmd	lame duck --> predicted: sequence-specific DNA binding transcription factor activity	myoblast fusion; somatic muscle development; skeletal muscle tissue development; muscle organ development	pharyngeal muscle; embryonic muscle system; embryonic somatic muscle; embryonic myoblast; embryonic visceral muscle	highest(high) 6-10hr embryo, mod until 14 hr embryo, very low or low for the rest except for white prepupa 24 hr where it is mod	x2	NB	NB	NB	NB	NB
21	CG6765	predicted: zinc ion binding, zinc finger like domain	no data	no data	Peak expression observed within 12-24 hour embryonic stages, during early larval stages	2/5	NB	NB	NB	NB	NB
22	stwl	stonewall --> predicted: DNA binding	germ-line stem cell maintenance; chromatin organization	ovary; germline cell	highest(mod high) during 0-10hr embryo, mod during 10-16 hr embryo, L3 through 2d pupa, 5d old adult male, adult female, and the rest is low.	x2	NB	NB	NB	NB	NB

Table 3-2 Gene Ontology characteristics of transcription factors identified by yeast-one-hybrid to bind to *yellow* enhancers

	Transcription factor that putatively binds to *yellow* enhancers	molecular function	biological process it is involved in	mutant phenotypes annotated with	expression	mel 5'	mel intron	pse 5'	pse intron	will 5'	will intron
23	chn	charlatan --> protein binding; sequence-specific DNA binding; sequence-specific enhancer binding RNA polymerase II transcription factor activity	progression of morphogenetic furrow involved in compound eye morphogenesis; eye development; positive regulation of transcription from RNA polymerase II promoter; sensory organ development; embryonic development via the syncytial blastoderm; negative regulation of compound eye cone cell fate specification; peripheral nervous system development	adult segment; organ system subdivision; peripheral nervous system; organ system; external compound sense organ; integumentary plate; postalar bristle; imaginal precursor; embryonic heart; epithelial furrow	highest(high) during 0-8hr embryo, mod high during white prepupa and 2d pupae as well as rest of the embryonic stages, moderate for the rest	x2	NB	NB	x1	NB	NB
24	Jra	Jun-related antigen --> protein binding	negative regulation of antimicrobial humoral response; micropyle formation; JNK cascade; synaptic growth at neuromuscular junction; imaginal disc fusion, thorax closure; dorsal appendage formation; phagocytosis, engulfment.	external compound sense organ; organ system subdivision; peripheral nervous system; nervous system; imaginal precursor; rhabdomere; primordium; commissure; cell projection; ommatidial precursor cluster	highest (high) at 18-20 hr embryo, and new white prepupa, mod high for the rest	NB	NB	x2	NB	NB	NB
25	ind	intermediate neuroblasts defective --> sequence-specific DNA binding transcription factor activity	regulation of transcription from RNA polymerase II promoter; negative regulation of transcription from RNA polymerase II promoter	RP2 neuron; neuroblast of ventral nerve cord primordium; intermediate ventral neurectoderm	highest (mod high) at 4-6-hr embryo, very low or non-existent for the rest	NB	NB	NB	NB	x2	NB
26	bigmax	sequence-specific DNA binding transcription factor activity	dendrite morphogenesis; muscle organ development; autophagic cell death; salivary gland cell autophagic cell death	no data	highest(high) during 18-24-hr embryo and white prepupae 12 hr, mod high for the rest	NB	X1	NB	NB	x2	NB
27	CrebB-17A	Cyclic-AMP response element binding protein B at 17A --> sequence-specific DNA binding	regulation of transcription, DNA-dependent; positive regulation of transcription from RNA polymerase II promoter; long-term memory; medium-term memory; sleep	adult; cuprophilic cell; synapse	mod high during embryo and adult, mod for the rest	NB	x1	NB	x1	NB	NB
28	gl	glass --> predicted: sequence-specific DNA binding transcription factor activity	entrainment of circadian clock; response to red light; ring gland development; compound eye photoreceptor fate commitment; entrainment of circadian clock by photoperiod	organ system subdivision; peripheral nervous system; nervous system; external compound sense organ; multi-cell-component structure; adult segment; endocrine system; late extended germ band embryo; primordium; adult brain	highest(low) larva through adult male, very low for the rest	NB	NB	NB	x2	NB	NB
29	CG7928	predicted: zinc ion binding; nucleic acid binding	no data	no data	highest(high) at 2-4-hr embryo, mod high for the rest of embryo, mod for the rest	NB	NB	x1	NB	NB	x2
30	Hey	Hairy/E(spl)-related with YRPW motif --> predicted: sequence-specific DNA binding transcription factor activity	negative regulation of Notch signaling pathway (predicted)	no data	highest(mod) during 6-16 hr embryo, vry low for the rest	NB	NB	NB	NB	NB	x2
31	CG2052	predicted: zinc ion binding; nucleic acid binding	no data	no data	no data	x1	x1	NB	NB	NB	NB
32	CG8359	predicted: DNA binding	no data	no data	highest(mod high) during 6-12 hr embryo, moderate for most of the rest, low during 3 and 4 day old pupa	x1	x1	NB	NB	NB	NB
33	vis	vismay --> predicted: sequence-specific DNA binding transcription factor activity; transcription corepressor activity	positive regulation of transcription, DNA-dependent; spermatogenesis	no data	highest (low-mod) during L3 through adult male, low for the rest and adult female	x1	x1	NB	NB	NB	NB
34	slp2	sloppy paired 2 --> sequence-specific DNA binding transcription factor activity	regulation of transcription, DNA-dependent	abdominal segment 4; abdominal segment 1; embryonic/larval dorsal vessel; cuticle; abdominal segment 2; abdominal segment 3; denticle belt	highest(mod high) during 6-10 hr embryo, low to mod for the rest of the embryo, low L3 through 2d pupa, very low for the rest	x1	NB	NB	x1	NB	NB

Table 3-2 Gene Ontology characteristics of transcription factors identified by yeast-one-hybrid to bind to *yellow* enhancers

	Transcription factor that putatively binds to *yellow* enhancers	molecular function	biological process it is involved in	mutant phenotypes annotated with	expression	mel 5'	mel intron	pse 5'	pse intron	will 5'	will intron
35	CG7101	predicted: zinc finger like domain, DNA binding	no data	no data	highest (mod high) during 0-2hr embryo and adult fem, moderate during 2-18hr embryo and white prepupae, low for the rest	x1	NB	NB	NB	NB	x1
36	dac	dachshund --> protein binding	axon guidance; compound eye photoreceptor development; compound eye development; neuron differentiation; mushroom body development; photoreceptor cell fate specification; antennal joint development; genital disc sexually dimorphic development	organ system subdivision; adult segment; external compound sense organ; peripheral nervous system; synaptic neuropil subdomain; multi-cell-component structure; metathoracic metatarsus; antennal segment; metatarsus; imaginal precursor	highest(mod) during 10-24hr embryo and L3 through 2d pupae, low to very low for the rest.	x1	NB	NB	NB	NB	x1
37	CG31392	no data	no data	no data	no data	NB	NB	NB	NB	x1	x1
38	CG33221	predicted: zinc ion binding, nucleic acid binding	no data	no data	highest (very low) during L3 through adult male, extremely low for the rest and adlut fem	NB	x1	NB	NB	x1	x1
39	CG18619	basic leucine zipper transcription factor	regulation of transcription, DNA-dependent	no data	highest(high) during embryo and L3 through end of white prepupa, moderate to mod high for the rest	NB	NB	NB	NB	x1	x1
40	dsx	doublesex --> protein binding; protein homodimerization activity, DNA binding	imaginal disc-derived female genitalia development; imaginal disc-derived male genitalia development; genital disc development; axon midline choice point recognition	organ system subdivision; organ system; adult segment; gland; imaginal precursor; portion of tissue; hub cell; larval abdominal segment 8; nervous system; external compound sense organ; female germline cyst	Peak expression (moderate) observedin early pupal and in adult male stages. Moderate expression throughout the rest of the pupa.	NB	NB	NB	NB	x2	NB
41	nau	nautilus --> DNA binding; sequence-specific DNA binding transcription factor activity	muscle organ development	myoblast; egg; abdominal dorsal oblique muscle 4; midgut constriction; somatic mesoderm; embryonic muscle system; abdominal 1 dorsal acute muscle 3; egg chamber; cardioblast; embryonic/larval somatic muscle	highest(mod) 6-16hr embryo, low for the rest of the embryo, very low until 2d pupae, low2d pupae through adult male, very low in adult fem	NB	NB	NB	x1	x1	NB
42	foxo	forkhead box, sub-group O --> sequence-specific DNA binding transcription factor activity	biological regulation; response to stress; regulation of insulin receptor signaling pathway; cellular process; regulation of growth; response to DNA damage stimulus; cellular component organization or biogenesis; response to bacterium; response to nutrient levels; determination of adult lifespan; sensory organ development; primary metabolic process	organ system subdivision; organ system; adult; external compound sense organ; adult segment; region of integument; thoracic segment; peripheral nervous system; ectoderm derivative; muscle cell; mushroom body neuroblast; portion of tissue	Peak expression observed within 00-06 and 18-24 hour embryonic stages, during late pupal stages	x1	NB	NB	NB	NB	NB
43	ftz-f1	ftz transcription factor 1 --> DNA binding, Nuclear hormone receptor, ligand-binding	olfactory behavior; pupation; metamorphosis; instar larval or pupal development; imaginal disc derived leg morphogenesis; periodic partitioning	organ system; abdominal ventral denticle belt; thoracic segment; cuticle; gland; organ system subdivision; portion of tissue; larval abdominal segment; larval thorax; sensillum; heart primordium; abdominal 5 ventral denticle belt; dorsal thoracic disc; parasegment 14; parasegment 2; cephalopharyngeal skeleton; spiracle; adult segment	Peak expression observed within 12-24 hour embryonic stages (very high). Expressed at moderately high level at the late pupal stage	x1	NB	NB	NB	NB	NB
44	mld	molting defective --> zinc ion binding; nucleic acid binding	determination of adult lifespan; positive regulation of circadian sleep/wake cycle, sleep; long-term memory; ecdysone biosynthetic process	trichogen cell; mesothoracic tergum; ring gland		x1	NB	NB	NB	NB	NB
45	Med	Medea --> protein binding; sequence-specific DNA binding	anatomical structure development; biological regulation; cellular process involved in reproduction; regulation of developmental process; post-embryonic appendage morphogenesis; regionalization; multicellular organismal reproductive process; post-embryonic organ morphogenesis; cell proliferation; cellular component organization or biogenesis; dorsal/ventral axis specification	organ system; organ system subdivision; adult segment; nervous system; germarium; synapse; adult mesothoracic segment; appendage segment; somatic cell; thoracic segment	Peak expression observed during early pupal stages	x1	NB	NB	NB	NB	NB

Table 3-2 Gene Ontology characteristics of transcription factors identified by yeast-one-hybrid to bind to *yellow* enhancers

	Transcription factor that putatively binds to *yellow* enhancers	molecular function	biological process it is involved in	mutant phenotypes annotated with	expression	mel 5'	mel intron	pse 5'	pse intron	will 5'	will intron
46	Snoo	Sno oncogene --> predicted: nucleotide binding	negative regulation of decapentaplegic signaling pathway; neuron development; negative regulation of transforming growth factor beta receptor signaling pathway	organ system subdivision; adult segment; organ system; thoracic segment; external compound sense organ; adult; imaginal precursor; embryonic/larval imaginal precursor; larval head segment; integumentary plate; peripheral nervous system		NB	NB	x1	NB	NB	NB
47	cas	castor --> DNA binding	central nervous system development; mushroom body development; post-embryonic development; neuroblast development; negative regulation of transcription, DNA-dependent	organ system subdivision; nervous system; multi-cell-component structure; synaptic neuropil subdomain; adult segment; sensillum; somatic cell; peripheral nervous system; organ system; sense organ	Peak expression observed within 06-18 hour embryonic stages, very low during pupa	NB	NB	NB	NB	x1	NB
48	vvl	ventral veins lacking --> sequence-specific DNA binding transcription factor activity	motor axon guidance; brain development; brain segmentation; peripheral nervous system development.	organ system subdivision; organ system; adult segment; larval abdominal segment; non-connected developing system; sensory cluster; portion of tissue; adult; peripheral nervous system; region of integument	Peak expression (moderately high) observed within 06-18 hour embryonic stages, moderate during pupa	NB	NB	NB	NB	x1	NB
49	E(bx)	Enhancer of bithorax --> ligand-dependent nuclear receptor binding	cellular component organization or biogenesis; biological regulation; chromatin organization; organelle organization; anatomical structure development; chromosome organization; cellular component organization or biogenesis at cellular level; cell-cell signaling; gene expression; prepupal development	dendrite; polytene chromosome; mesothoracic tergum; melanotic mass	Peak expression (moderately high to high) observed within 00-12 hour embryonic stages, moderate to moderately high expression during pupa	NB	NB	NB	NB	x1	NB
50	fru	fruitless --> sequence-specific DNA binding transcription factor activity	mating; behavioral interaction between organisms; anatomical structure development; multicellular organismal reproductive process; multi-organism process; cellular component organization or biogenesis; male courtship behavior, veined wing extension; developmental process involved in reproduction; muscle organ development; central nervous system development	organ system subdivision; adult segment; synaptic neuropil subdomain; embryonic/larval neuron; nervous system; antennal segment; adult mesothoracic segment; embryonic/larval glial cell; presumptive embryonic/larval nervous system; gland; external compound sense organ	Peak expression observed within 18-24 hour embryonic stages, during early larval stages, during late pupal stages, in adult male stages	NB	NB	NB	NB	x1	NB
51	cyc	cycle --> protein heterodimerization activity; DNA binding	response to starvation; regulation of circadian sleep/wake cycle, sleep; circadian regulation of gene expression	LN period neuron	Peak expression observed within 00-06 hour embryonic stages, in adult female stages, moderate to moderately high during pupa	NB	x1	NB	NB	NB	NB
52	crm	cramped --> DNA binding	segment specification	adult segment; organ system subdivision; antennal segment; appendage segment; somatic cell; external compound sense organ; adult mesothoracic segment; thoracic segment; organ system; prothoracic leg; compound cell cluster organ; pigment cell; metatarsus	Peak expression observed within 00-06 hour embryonic stages, in adult female stages	NB	x1	NB	NB	NB	NB
53	gt	giant --> sequence-specific DNA binding transcription factor activity	torso signaling pathway; terminal region determination; specification of segmental identity, labial segment; ring gland development; axon guidance; regulation of gene expression; regulation of cell size; negative regulation of multicellular organism growth; phagocytosis, engulfment	organ system; abdominal segment 7; abdominal ventral denticle belt; embryonic segment; thoracic segment; embryonic abdomen; organ system subdivision; larval abdominal segment; multicellular structure; cephalopharyngeal skeleton; external compound sense organ; late extended germ band embryo	Peak expression observed within 00-06 hour embryonic stages, low expression during pupa	NB	x1	NB	NB	NB	NB

Table 3-2. Gene Ontology characteristics of transcription factors identified by yeast-one-hybrid to bind to *yellow* enhancers

Gene Ontology characterizations of transcription factors that were shown to interact with more than one *yellow* enhancer sub-element and/or were knocked-down using RNA interference in *D. melanogaster*. The 5' intergenic or intronic regions of *yellow* from three Drosophila species and the number of sub-elements from each that interactedwith the corresponding transcription factor is indicated in columns on the right side (e.g., x1 interacted with only one sub-element, 4/5 --> interacted with four out of five subelements in the particular region, NB: not binding)

Table 3-3 List of fly lines harboring UAS-RNA interference transgenes used to knock down *Drosophila melanogaster* transcription factors

BL#	TRiP#	CG#	FBgn	Gene Name	Vector	Genotype	Hairpin Location	Description	RNAi Expressed	Hairpin ID
31990	JF03424	CG7199	FBgn0015239	Hormone-receptor-like in 78	VALIUM10	y1 v1; P{TRiP}attP2	attP2		soma	TR01457P.1
29376	JF02540	CG1864	FBgn0014859	Hormone receptor-like in 38	VALIUM10	y1 v1; P{TRiP}attP2	attP2		soma	TR01451P.1
25845	JF01866	CG8318	FBgn0015269	Neurofibromin 1	VALIUM10	y1 v1; P{TRiP}attP2	attP2		soma	TR01730P.1
26718	JF02258	CG18023	FBgn0004865	Ecdysone-induced protein 78C	VALIUM10	y1 sc v1; P{TRiP}attP2	attP2		soma	TR01443P.1
31975	JF03409	CR4575	FBgn0029920	-	VALIUM10	y1 v1; P{TRiP}attP2/TM3,Sb	attP2	TM3,Sb	soma	TR02615P.1
27564	JF02718	CG6376	FBgn0011766	E2F transcription factor	VALIUM10	y1 v1; P{TRiP}attP2	attP2		soma	TR00300P.1
26746	JF02309	CG11648	FBgn0000015	Abdominal B	VALIUM10	y1 v1; P{TRiP}attP2	attP2		soma	TR01038P.1
27315	JF02624	CG5460	FBgn0001169	Hairless	VALIUM10	y1 v1; P{TRiP}attP2/TM3,Sb	attP2	TM3,Sb	soma	TR02167P.1
34703	HMS01182	CG5460	FBgn0001169	Hairless	VALIUM20	y1 sc v1; P{TRiP}attP2	attP2		soma, gern	SH01852.N
29355	JF02518	CG6312	FBgn0020379	Rfx	VALIUM10	y1 v1; P{TRiP}attP2	attP2		soma	TR01264P.1
25923	JF01943	CG42702	FBgn0261613	O/E-associated zinc finger protein	VALIUM10	y1 v1; P{TRiP}attP2	attP2		soma	TR01862P.1
29407	JF03343	CG4807	FBgn0259750	abrupt	VALIUM10	y1 v1; P{TRiP}attP2	attP2		soma	TR02394P.1
28338	JF02973	CG34395	FBgn0085424	nubbin	VALIUM10	y1 v1; P{TRiP}attP2	attP2		soma	TR01098P.1
33699	HMS00574	CG11033	FBgn0037659	-	VALIUM20	y1 sc v1; P{TRiP}attP2	attP2		soma, gern	SH00815.N
26779	JF02343	CG11798	FBgn0015371	charlatan	VALIUM10	y1 v1; P{TRiP}attP2/TM3,Sb	attP2	TM3,Sb	soma	TR01405P.1
31595	JF01184	CG2275	FBgn0001291	Jun-related antigen	VALIUM1	y1 v1; P{TRiP}attP2	attP2		soma	TR00327A.1
28662	JF03077	CG11551	FBgn0025776	intermediate neuroblasts defective	VALIUM10	y1 v1; P{TRiP}attP2	attP2		soma	TR02650P.1
29325	JF02487	CG3350	FBgn0039509	bigmax	VALIUM10	y1 v1; P{TRiP}attP2	attP2		soma	TR00883P.1
29332	JF02494	CG6103	FBgn0014467	Cyclic-AMP response element binding protein B at 17A	VALIUM10	y1 v1; P{TRiP}attP2/TM3,Sb	attP2	TM3,Sb	soma	TR00977P.1
26780	JF02344	CG7672	FBgn0004618	glass	VALIUM10	y1 v1; P{TRiP}attP2	attP2		soma	TR01409P.1
31898	JF02186	CG11194	FBgn0027788	Hairy/E(spl)-related with YRPW motif	VALIUM10	y1 v1; P{TRiP}attP2	attP2		soma	TR00912P.1
26711	JF02251	CG2052	FBgn0262636	CG2052	VALIUM10	y1 v1; P{TRiP}attP2	attP2		soma	TR01389P.1
27066	JF02411	CG8359	FBgn0037634	CG8359	VALIUM10	y1 v1; P{TRiP}attP2	attP2		soma	TR01176P.1
29544	JF02406	CG8821	FBgn0033748	vismay	VALIUM10	y1 v1; P{TRiP}attP2	attP2		soma	TR01122P.1
34634	HMS01108	CG2939	FBgn0004567	sloppy paired 2	VALIUM20	y1 sc v1; P{TRiP}attP2	attP2		soma, gern	SH01863.N
26758	JF02322	CG4952	FBgn0005677	dachshund	VALIUM10	y1 v1; P{TRiP}attP2	attP2		soma	TR01227P.1
29542	JF02401	CG18619	FBgn0032202	-	VALIUM10	y1 v1; P{TRiP}attP2	attP2		soma	TR00967P.1
31899	JF02187	CG10250	FBgn0002922	nautilus	VALIUM10	y1 v1; P{TRiP}attP2/TM3,Sb	attP2	TM3,Sb	soma	TR00931P.1
25997	JF02019	CG3143	FBgn0038197	forkhead box, sub-group O	VALIUM10	y1 v1; P{TRiP}attP2	attP2		soma	TR01256P.1
31928	JF02218	CG1775	FBgn0011655	Medea	VALIUM10	y1 v1; P{TRiP}attP2	attP2		soma	TR01193P.1
31593	JF01182	CG14307	FBgn0004652	fruitless	VALIUM1	y1 v1; P{TRiP}attP2	attP2		soma	TR00308A.1
26228	JF02126	CG10037	FBgn0086680	ventral veins lacking	VALIUM10	y1 v1; P{TRiP}attP2	attP2		soma	TR01123P.1
33658	HMS00065	CG32346	FBgn0000541	Enhancer of bithorax	VALIUM20	y1 sc v1; P{TRiP}attP2	attP2		soma, gern	SH00335.N
26310	JF02083	CG2102	FBgn0004878	castor	VALIUM10	y1 v1; P{TRiP}attP2	attP2		soma	TR00856P.1
31952	JF02244	CG34100	FBgn0083077	molting defective	VALIUM10	y1 v1; P{TRiP}attP2	attP2		soma	TR01354P.1
31934	JF02225	CG34421	FBgn0085450	Sno oncogene	VALIUM10	y1 v1; P{TRiP}attP2	attP2		soma	TR01229P.1
26716	JF02256	CG11094	FBgn0000504	doublesex	VALIUM10	y1 v1; P{TRiP}attP2	attP2		soma	TR01431P.1
26742	JF02304	CG7952	FBgn0001150	giant	VALIUM10	y1 v1; P{TRiP}attP2	attP2		soma	TR00978P.1
27659	JF02738	CG4059	FBgn0001078	ftz transcription factor 1	VALIUM10	y1 v1; P{TRiP}attP2	attP2		soma	TR01448P.1
29400	JF03333	CG8727	FBgn0023094	cycle	VALIUM10	y1 v1; P{TRiP}attP2	attP2		soma	TR00889P.1
31338	JF01296	CG2714	FBgn0000376	cramped	VALIUM1	y1 v1; P{TRiP}attP2/TM3,Ser	attP2	TM3,Ser	soma	TR00749A.1

BL: Bloomington Stock Center
TRIP: Transgenic RNAi Project
FBgn: Flybase Gene Number

References

Andrews, H. K., N. Giagtzoglou, S. Yamamoto, K. L. Schulze, and H. J. Bellen. 2009. Sequoia regulates cell fate decisions in the external sensory organs of adult Drosophila. EMBO reports 10:636-641.

Baker, B. S., and K. A. Ridge. 1980. Sex and the single cell. I. On the action of major loci affecting sex determination in Drosophila melanogaster. Genetics 94:383-423.

Baker, B. S., B. J. Taylor, and J. C. Hall. 2001. Are complex behaviors specified by dedicated regulatory genes? Reasoning from Drosophila. Cell 105:13-24.

Brenman, J. E., F. B. Gao, L. Y. Jan, and Y. N. Jan. 2001. Sequoia, a tramtrack-related zinc finger protein, functions as a pan-neural regulator for dendrite and axon morphogenesis in Drosophila. Developmental Cell 1:667-677.

Britten, R. J., and E. H. Davidson. 1971. Repetitive and non-repetitive DNA sequences and a speculation on the origins of evolutionary novelty. The Quarterly review of biology 46:111-138.

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

Cléard, F., and P. Spierer. 2001. Position-effect variegation in Drosophila: the modifier Su(var)3-7 is a modular DNA-binding protein. EMBO reports 2:1095-1100.

Couderc, J. L., D. Godt, S. Zollman, J. Chen, M. Li, S. Tiong, S. E. Cramton, I. Sahut-Barnola, and F. A. Laski. 2002. The bric a brac locus consists of two paralogous genes encoding BTB/POZ domain proteins and acts as a homeotic and morphogenetic regulator of imaginal development in Drosophila. Development 129:2419-2433.

de Celis, J. F., M. Llimargas, and J. Casanova. 1995. Ventral veinless, the gene encoding the Cf1a transcription factor, links positional information and cell differentiation during embryonic and imaginal development in Drosophila melanogaster. Development 121:3405-3416.

de Hoon, M. J. L., S. Imoto, J. Nolan, and S. Miyano. 2004. Open source clustering software. Bioinformatics 20:1453-1454.

Diaz-Benjumea, J., M. A. F. Gonzalez Gaitan, and A. Garcia-Bellido. 1989. Developmental genetics of the wing vein pattern of *Drosophila*. Genome 31:612-619.

Drapeau, M. 2006. FRU does not regulate yellow in fly pigmented cells. PHILICA.COM Observation number 13.

Drapeau, M. D., S. A. Cyran, M. M. Viering, P. K. Geyer, and A. D. Long. 2005. A cis-regulatory Sequence Within the yellow Locus of Drosophila melanogaster Required for Normal Male Mating Success. Genetics 172:1009-1030.

Drapeau, M. D., A. Radovic, P. J. Wittkopp, and A. D. Long. 2003. A gene necessary for normal male courtship, yellow, acts downstream of *fruitless* in the *Drosophila melanogaster* larval brain. Journal of Neurobiology 55:53-72.

Dubruille, R., A. Laurençon, C. Vandaele, E. Shishido, M. Coulon-Bublex, P. Swoboda, P. Couble, M. Kernan, and B. Durand. 2002. Drosophila Regulatory factor X is necessary for ciliated sensory neuron differentiation. Development 129:5487-5498.

Durand, B., C. Vandaele, D. Spencer, S. Pantalacci, and P. Couble. 2000. Cloning and characterization of dRFX, the Drosophila member of the RFX family of transcription factors. Gene 246:285-293.

Eichenlaub, M. P., and L. Ettwiller. 2011. De Novo Genesis of Enhancers in Vertebrates. PLoS Biology 9:e1001188.

Fisk, G. J., and C. S. Thummel. 1998. The DHR78 Nuclear Receptor Is Required for Ecdysteroid Signaling during the Onset of Drosophila Metamorphosis. Cell 93:543-555.

Frankel, N., G. K. Davis, D. Vargas, S. Wang, F. Payre, and D. L. Stern. 2010. Phenotypic robustness conferred by apparently redundant transcriptional enhancers. Nature 466:490-493.

Frankel, N., D. F. Erezyilmaz, A. P. McGregor, S. Wang, F. Payre, and D. L. Stern. 2011. Morphological evolution caused by many subtle-effect substitutions in regulatory DNA. Nature 474:598-603.

Futahashi, R., and H. Fujiwara. 2007. Regulation of 20-hydroxyecdysone on the larval pigmentation and the expression of melanin synthesis enzymes and yellow gene of the swallowtail butterfly, Papilio xuthus. Insect Biochemistry and Molecular Biology 37:855-864.

Geyer, P. K., and V. G. Corces. 1987. Separate regulatory elements are responsible for the complex pattern of tissue-specific and developmental transcription of the yellow locus in Drosophila melanogaster. Genes & Development 1:996-1004.

Guo, H. F., J. Tong, F. Hannan, L. Luo, and Y. Zhong. 2000. A neurofibromatosis-1-regulated pathway is required for learning in Drosophila. Nature 403:895-898.

Hare, E. E., B. K. Peterson, V. N. Iyer, R. Meier, and M. B. Eisen. 2008. Sepsid even-skipped Enhancers Are Functionally Conserved in Drosophila Despite Lack of Sequence Conservation. PLoS Genetics 4:e1000106.

Heitzler, P., M. Haenlin, P. Ramain, M. Calleja, and P. Simpson. 1996. A Genetic Analysis of pannier, a Gene Necessary for Viability of Dorsal Tissues and Bristle Positioning in Drosophila. Genetics 143:1271-1286.

Hens, K., J. D. Feuz, A. Isakova, A. Iagovitina, A. Massouras, J. Bryois, P. Callaerts, S. E. Celniker, and B. Deplancke. 2011. Automated protein-DNA interaction screening of Drosophila regulatory elements. Nat Methods 8:1065-1070.

Ho, I. S., F. Hannan, H. F. Guo, I. Hakker, and Y. Zhong. 2007. Distinct Functional Domains of Neurofibromatosis Type 1 Regulate Immediate versus Long-Term Memory Formation. Journal of Neuroscience 27:6852-6857.

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

Jeong, S., A. Rokas, and S. B. Carroll. 2006. Regulation of Body Pigmentation by the Abdominal-B Hox Protein and Its Gain and Loss in Drosophila Evolution. Cell 125:1387-1399.

Kopp, A., R. K. Blackman, and I. Duncan. 1999. Wingless, Decapentaplegic and EGF Receptor signaling pathways interact to specify dorso-ventral pattern in the adult abdomen of Drosophila. Development 126:3495-3507.

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

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

Kozlova, T., G. Lam, and C. S. Thummel. 2009. DrosophilaDHR38 nuclear receptor is required for adult cuticle integrity at eclosion. Developmental Dynamics 238:701-707.

Lee, M.-J., and D. A. Stephenson. 2007. Recent developments in neurofibromatosis type 1. Current opinion in neurology 20:135-141.

Llimargas, M., and J. Casanova. 1997. ventral veinless, a POU domain transcription factor, regulates different transduction pathways required for tracheal branching in Drosophila. Development 124:3273-3281.

Martin, M., Y. B. Meng, and W. Chia. 1989. Regulatory elements involved in the tissue-specific expression of the yellow gene of Drosophila. Mol Gen Genet 218:118-126.

McGregor, A. P., V. Orgogozo, I. Delon, J. Zanet, D. G. Srinivasan, F. Payre, and D. L. Stern. 2007. Morphological evolution through multiple cis-regulatory mutations at a single gene. Nature 448:587-590.

Meier, S., S. G. Sprecher, H. Reichert, and F. Hirth. 2006. ventral veins lacking is required for specification of the tritocerebrum in embryonic brain development of Drosophila. Mech Dev 123:76-83.

Ni, J.-Q., M. Markstein, R. Binari, B. Pfeiffer, L.-P. Liu, C. Villalta, M. Booker, L. Perkins, and N. Perrimon. 2008. Vector and parameters for targeted transgenic RNA interference in Drosophila melanogaster. Nature Methods 5:49-51.

Ni, J. Q., L. P. Liu, R. Binari, R. Hardy, H. S. Shim, A. Cavallaro, M. Booker, B. D. Pfeiffer, M. Markstein, H. Wang, C. Villalta, T. R. Laverty, L. A. Perkins, and N. Perrimon. 2009. A Drosophila Resource of Transgenic RNAi Lines for Neurogenetics. Genetics 182:1089-1100.

Pan, D., S. A. Valentine, and A. J. Courey. 1994. The bipartite D. melanogaster twist promoter is reorganized in D. virilis. Mechanisms of development 46:41-53.

Pospisilik, J. A., D. Schramek, H. Schnidar, S. J. F. Cronin, N. T. Nehme, X. Zhang, C. Knauf, P. D. Cani, K. Aumayr, J. Todoric, M. Bayer, A. Haschemi, V. Puviindran, K. Tar, M. Orthofer, G. G. Neely, G. Dietzl, A. Manoukian, M. Funovics, G. Prager, O. Wagner, D. Ferrandon, F. Aberger, C.-c. Hui, H. Esterbauer, and J. M. Penninger. 2010. Drosophila Genome-wide Obesity Screen Reveals Hedgehog as a Determinant of Brown versus White Adipose Cell Fate. Cell 140:148-160.

Prabhakar, S., A. Visel, J. A. Akiyama, M. Shoukry, K. D. Lewis, A. Holt, I. Plajzer-Frick, H. Morrison, D. R. FitzPatrick, V. Afzal, L. A. Pennacchio, E. M. Rubin, and J. P. Noonan. 2008. Human-specific gain of function in a developmental enhancer. Science 321:1346-1350.

Rebeiz, M., N. Jikomes, V. Kassner, and S. Carroll. 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 108:10036-10043.

Reuter, G., M. Giarre, J. Farah, J. Gausz, A. Spierer, and P. Spierer. 1990. Dependence of position-effect variegation in Drosophila on dose of a gene encoding an unusual zinc-finger protein. Nature 344:219-223.

Russell, S. R., G. Heimbeck, C. M. Goddard, A. T. Carpenter, and M. Ashburner. 1996. The Drosophila Eip78C gene is not vital but has a role in regulating chromosome puffs. Genetics 144:159-170.

Sanges, R., E. Kalmar, P. Claudiani, M. D'Amato, F. Muller, and E. Stupka. 2006. Shuffling of cis-regulatory elements is a pervasive feature of the vertebrate lineage. Genome Biology 7:R56.

Sotillos, S., and J. F. de Celis. 2006. Regulation of decapentaplegic expression during Drosophila wing veins pupal development. Mech Dev 123:241-251.

Spierer, A., F. Begeot, P. Spierer, and M. Delattre. 2008. SU(VAR)3-7 Links Heterochromatin and Dosage Compensation in Drosophila. PLoS Genetics 4:e1000066.

Stone, J. R., and G. A. Wray. 2001. Rapid evolution of cis-regulatory sequences via local point mutations. Molecular Biology and Evolution 18:1764-1770.

Sumiyama, K., and N. Saitou. 2011. Loss-of-Function Mutation in a Repressor Module of Human-Specifically Activated Enhancer HACNS1. Molecular Biology and Evolution 28:3005-3007.

Swanson, C. I., N. C. Evans, and S. Barolo. 2010. Structural rules and complex regulatory circuitry constrain expression of a Notch- and EGFR-regulated eye enhancer. Developmental Cell 18:359-370.

The, I., G. Hannigan, G. Cowley, S. Reginald, Y. Zhong, J. Gusella, I. Hariharan, and A. Bernards. 1997. Rescue of a Drosophila NF1 Mutant Phenotype by Protein Kinase A. Science 276:791-794.

Tong, J. J., S. E. Schriner, D. McCleary, B. J. Day, and D. C. Wallace. 2007. Life extension through neurofibromin mitochondrial regulation and antioxidant therapy for neurofibromatosis-1 in Drosophila melanogaster. Nature Genetics 39:476-485.

Vandaele, C., M. Coulon-Bublex, P. Couble, and B. Durand. 2001. Drosophila regulatory factor X is an embryonic type I sensory neuron marker also expressed in spermatids and in the brain of Drosophila. Mechanisms of development 103:159-162.

Werner, T., S. Koshikawa, T. M. Williams, and S. B. Carroll. 2010. Generation of a novel wing colour pattern by the Wingless morphogen. Nature 464:1143-1148.

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

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

Wittkopp, P. J., J. R. True, and S. B. Carroll. 2002a. Reciprocal functions of the Drosophila Yellow and Ebony proteins in the development and evolution of pigment patterns. Development 129:1849-1858.

Wittkopp, P. J., K. Vaccaro, and S. B. Carroll. 2002b. Evolution of yellow Gene Regulation and Pigmentation in Drosophila. Current Biology 12:1547-1556.

Yuh, C. H., and E. H. Davidson. 1996. Modular cis-regulatory organization of Endo16, a gut-specific gene of the sea urchin embryo. Development 122:1069-1082.

Yuh, C. H., J. G. Moore, and E. H. Davidson. 1996. Quantitative functional interrelations within the cis-regulatory system of the S. purpuratus Endo16 gene. Development 122:4045-4056.

Chapter 4

Conclusion

Gene expression is a crucial step on the way from DNA to the ultimate gene product. Changes in gene expression patterns during development can lead to aberrant phenotypes (Schneuwly et al. 1987), but for evolution these changes have been shown to be a major source of phenotypic diversity (Carroll 2008). Gene expression is controlled by *cis* and *trans* regulatory factors. *cis*-regulatory elements are linked to the gene they affect whereas *trans*-regulatory factors are typically diffusible molecules, unlinked to the gene(s) they affect. In order to elucidate how regulation of gene expression works and evolves, one needs to understand the changes in both *cis* and *trans* regulatory factors as well as the relationship between the two. In this thesis, I focused on the changes in *cis*-regulatory elements, particularly enhancers, and studied them using functional *in vivo* assays combined with sequence comparisons and *in vitro* analyses. As a model, I used enhancers regulating the rapidly diverging expression patterns of the Drosophila *yellow* gene.

I first set out to identify three enhancers of *yellow* that drive the gene's expression in the body, wings and bristle associated cells from six species (*D. melanogaster, D. pseudoobscura, D. willistoni, D. mojavensis, D. virilis, D. grimshawi*) that span the Drosophila evolutionary history. In order to assess if the genomic organization of *yellow* enhancers was conserved over evolutionary time, in each species I tested *yellow* 5’ intergenic and intronic regions of *yellow*, where the above-mentioned three enhancers are typically located, for enhancer activity. I found that the location of body and wing enhancers are highly variable among Drosophila species, such that several species have these enhancer activities in both regions tested, whereas in some species only one of the tested regions have body and wing enhancer activities. Overall, this suggests that the positions of *yellow* body and wing enhancer activities were altered, with respect to the coding sequence, multiple times in the evolutionary history. Most intriguingly, these results showed, against conventional wisdom, that enhancer position is not always conserved among species, and in fact it can be quite labile. They also showed that in all but one species (*D. grimshawi*), *yellow* body and wing enhancers were in the same relative genomic position, suggesting these two enhancers are sharing a large number of transcription factor binding sites and perhaps can be referred to as a single enhancer called the “epidermal-cell” enhancer. This recategorized enhancers of the *yellow* gene as “epidermal-cell”, “wing-vein”, and “bristle”. Over evolutionary time, the first two appeared to have changed positions independent of each other, whereas the bristle enhancer was stably located in the intron. Among these three enhancers, I also found that the spatial expression patterns driven by the epidermal-cell enhancer was highly variable among species. Lastly, sequence comparisons of *yellow* enhancers between closely related Drosophila species did not show signs of any large scale genomic rearrangements suggesting that *yellow* epidermal-cell and wing-vein enhancer activities changed position through gradual compensatory sequence changes.

Even though the above findings addressed important questions (e.g., what are the activities and genomic positions of *yellow* enhancers and are they conserved among species?) identifying how activity and position of *yellow* enhancers changed over evolutionary time requires more detailed functional analysis of these enhancers as well as comparisons among them. In order to achieve that I first subdivided the previously tested *yellow* 5’ intergenic and intronic regions from three Sophophora subgroup species, *D. melanogaster*, *D. pseudoobscura*, *D. willistoni*, and tested the resulting sub-elements for enhancer activity. This study revealed sub-elements that fully or partially recapitulate the expression pattern driven by the full region they are isolated from as well as sub-elements that do not drive expression above background levels. In addition, and most intriguingly, I found that about half of the sub-elements tested harbored cryptic enhancer activities, such that the sub-element drove expression in spatial patterns that were not part of the expression pattern driven by the full 5' intergenic or intronic region the sub-element was isolated from. This may be an effect of the change in the proximity of the particular sub-element to the transcription start site when being tested for enhancer activity or the effect of the foreign genomic location where all transgenes were tested. Both of these are viable possibilities that need to be tested, however, at the same time, they are unlikely possibilities because a little less than one third of the sub-elements tested did not show enhancer activity despite change in proximity and genomic position. Thus, if there is a relationship, for example between the flanking sequences of the transgene insertion site and the sequences of some, but not all, sub-elements, this relationship is quite complex. It is therefore possible that the cryptic enhancer activities observed in certain (about half of the) sub-elements are real and repressed by the neighboring sequences when the sub-element is in its native position (i. e., in the *yellow* 5' intergenic or intronic regions that it is isolated from). Existence of cryptic activities in the 5' intergenic or intronic regions of *yellow* presents a possible scenario for how and why position and activity of *yellow* enhancers have diverged rapidly. That is, when a region is primed for a certain enhancer activity (i.e., harbors part of the necessary set transcription factor binding sites) it takes arguably fewer mutational steps to attain that enhancer activity than it would have taken if the same activity was created from non-functional sequence.

The next step in understanding how enhancers change over time is to identify the transcription factors that bind to them and subsequently shed light onto how changes in the binding factors of enhancers affect their activity. In order to identify the sets of transcription factors that bind to *yellow* enhancers, I screened the previously tested *yellow* enhancer sub-elements using yeast-one-hybrid (Y1H) with an extensive *D. melanogaster* transcription factor library (85% of all transcription factors, n=647)). As a result, I identified a large set of candidate transcription factors binding to *yellow* enhancers from three different Drosophila species. Only ~10% of these were shared between physically homologous regions between species even though these regions had similar enhancer activities (e.g., epidermal-cell enhancer). This is in part due to the false negative rate of Y1H, and further validation assays may show transcription factors that interacted with *yellow* enhancers in only one species are interactors of *yellow* enhancers in other Drosophila species as well. Some of the identified transcription factors suggest that *yellow* appears to be a target of the ecdysone signaling pathway, which, in flies, is responsible for the major postembryonic developmental transitions. It also appears to be regulated by genes important for bristle development, ones that are important for transcription from the X chromosome (where *yellow* is located) overall and even by a gene that causes tumors in humans. Preliminary validation assays, for the first time, identified two transcription factors, Fruitless and Ventral veins lacking, to affect abdominal pigmentation, potentially through regulating *yellow* expression. Functionally validating these candidate transcription factors identified by Y1H as real direct binders of *yellow* enhancers, will likely identify new pathways that *yellow* is a target or part of.

After identifying the transcription factors binding to *yellow* enhancers, in order to elucidate the relationship between transcription factor binding and enhancer activity, one can ask whether enhancers with similar activities are bound by the similar or dissimilar sets of transcription factors. Even though the enhancer activity and transcription factor binding profile datasets I acquired were not fully completed/validated, I still conducted a preliminary comparison between a qualitative assessment of the activities of *yellow* enhancer sub-elements and their transcription factor binding profiles. The only strong relationship I observed was the similarity of transcription factor binding profiles of *yellow* enhancer sub-elements that showed male specific abdominal enhancer activity (higher expression in male abdominal tergites A5 and A6 as compared to that of females). Comparing a quantitative assessment of the expression patterns driven by *yellow* enhancer sub-elements at different developmental stages to a more complete and validated list of transcription factors binding to these sub-elements can help come to a more general conclusion about whether enhancers with similar activities are bound by similar sets of transcription factors.

How common is enhancer position change?

One of the most intriguing results of this thesis is that enhancer position, with respect to the coding sequences, can be labile. This is the first time enhancer position was systematically analyzed among species and the results are against the general assumption that functionally homologous enhancers have conserved positions among species, i. e., if an enhancer is found at a certain position in one species, it is expected to be located in the physically homologous region in a related species. The commonality of this finding is not yet known. It is possible that *yellow* is unique because it is located at the tip of the X chromosome where recombination rate is lower and chances of accumulating deleterious mutations is higher than it is in other genomic regions. Also as compared to the early developmental genes, as a late developmental gene, mutations changing *yellow* expression patterns may be more tolerable for the organism and transitionary expression states may have a higher chance of surviving in nature until finding a new optimal state.

It is also possible, however, that enhancer position change is not unique to *yellow* and the reason for the rarity of its examples in the literature is the ascertainment bias towards not publishing negative results. For instance, if an enhancer is not found in the same position (relative to the coding sequence) as it was in another species this type of negative result is typically not documented. In that sense, the results of this thesis directly affect future searches for functionally homologous enhancers, by showing that they may be in different positions in different species. Forthcoming studies that conduct exhaustive analyses to find all regions harboring a certain type of enhancer activity contributing to a gene's full expression pattern (for instance at a particular tissue type) in different species can shed light onto the generality of enhancer position change.

Do redundant enhancers change position more readily?

It is important to note that in some Drosophila species epidermal cell enhancer activity is distributed between the 5' intergenic and intronic regions of *yellow* in a partially overlapping and partially complementary manner, such that some spatial patterns are driven by only one region, but some are driven by both regions. In the case of *D. virilis yellow*, for instance, both the 5' intergenic and intronic regions harbor similar enhancer activities such that (if pigmentation is taken as a proxy for *yellow* expression) either one could recapitulate the expression of the gene in the body epidermal cells. In fact, in Chapter 2, epidermal-cell and wing-vein enhancer activities are inferred to exist in both the 5' intergenic and intronic regions of *yellow* in the (hypothetical) common ancestor of Drosophila species, and perhaps this redundancy relaxed the selective constraint on *yellow* enhancers such that loss or change of activity in one region (e.g., 5' intergenic) was compensated by the other region (e.g. intron) harboring a similar enhancer activity. Bristle enhancer, however, was unique in the *yellow* intron and remained that way throughout Drosophila evolution, arguably because transitionary states towards changing its position and activity could not be compensated by a secondary bristle enhancer, and hence, such changes were selected against. This third feature may in fact not be unique to *yellow*, but may not have been revealed for other genes as it requires a comprehensive analysis of the *cis*-regulatory regions of a gene.

Interestingly, in recent years, studies showed that for a particular expression pattern, the *cis*-regulatory regions of an increasing number of genes harbor more enhancers than minimally necessary (Perry et al. 2010; Frankel et al. 2011; Perry et al. 2011). Among such enhancers, even though some appear “informationally” redundant (Barolo 2012) with each other, they in fact act synergistically to contribute to the robustness of the particular expression pattern, especially under stress conditions. These enhancers have so far been found in only one (Perry et al. 2010; Perry et al. 2011) or few species (Frankel et al. 2010a) that are closely related and among which the position of the redundant enhancers were conserved. There is one study, however, where the authors investigated conservation of location of Dorsal target enhancers among distantly related species (Cande et al. 2009a). Among the six genes they looked at, three (*short gastrulation* (*sog*), *brinker* (*brk*) and *ventral nervous system defective* (*vnd*)) harbored two informationally redundant Dorsal target enhancers in one species (*D. melanogaster*) but in the other species (*Anopheles gambiae* or *Tribolium castaneum*) one of the redundant enhancers was lost, indicating a similar case to *yellow* epidermal-cell enhancers. The other three genes (*cactus* (*cact*), *twist* (*twi*) and *single-minded* (*sim*)) appeared to harbor a single Dorsal target enhancer, the position of which was conserved among species, suggesting that there is more selective constraint on positions of “single” enhancers as compared to informationally redundant enhancers. It is notable, however, that among the latter three genes *twi* has a proximal and a distal enhancer that appear to drive overlapping expression patterns and the proximal enhancer has been found to be located in the intron of the gene in *D. virilis* as opposed to the 5‘ intergenic location in *D. melanogaster* (Pan et al. 1994a). This, once again suggests, that enhancer position change between species may be allowed by the partial redundancy between the two enhancers. With careful and unbiased experimental designs, elucidating positions of more of these “informationally” redundant enhancers between distantly related species (e.g., *D. melanogaster* and *D. pseudoobscura* or even further) can shed light onto whether redundant enhancers are more likely to change their positions over evolutionary time. At the experimental level this would require comprehensively testing all candidate regions that may harbor enhancer activity without stopping after finding one (minimal) enhancer. For instance, in the case of *D. virilis yellow* epidermal cell enhancer: if one were to find the 5’ enhancer, looking in the intron for more enhancer activity of the same type would seem unnecessary, which would result in incomplete knowledge of enhancers and how they affect gene expression. Future studies that conduct comprehensive analyses of *cis*-regulatory sequences can help us better understand the full structure and function of enhancers. Fortunately new methods for *in vivo* testing of candidate regions for enhancer activity in a high throughput manner (Pennacchio et al. 2006; Weiszmann et al. 2009) are up and coming.

Cryptic enhancer activities

Recent studies showed examples of cryptic enhancer activity that allowed different evolutionary trajectories, which resulted in altered gene expression patterns. These were initially thought to be new enhancer activities prior to the discovery of the latent activity. In one case a partial ancestral enhancer lost its activity fully due to repressive mutations in some lineages, but became a full enhancer in other lineages through gaining activating mutations and co-opting neighboring enhancers (Rebeiz et al. 2011b). In another case, disruption of ancestral repressive sequences revealed latent *cis*-regulatory activity (Sumiyama and Saitou 2011). As discussed in Chapter 3 and earlier in this chapter, if the cryptic enhancer activities observed in almost half of the tested *yellow* enhancer sub-elements are not experimental artifacts, then existence of these activities may have facilitated rapid change in the activity and position of *yellow* enhancers. This seems particularly evident in the epidermal-cell enhancer since all cryptic activities identified so far drove expression in body epidermal cells. That is to say, it is likely that the 5’ intergenic and intronic regions of *yellow* were ancestrally primed such that they harbored a collection of binding sites for appropriate transcription factors (e.g. that are expressed in a specific tissue at appropriate developmental stages). With the existence of such a “basis” for regulatory information in a DNA region, evolution of altered enhancer activities (in the specific tissue type) would require fewer mutations than it would if the region harbored putatively non-functional sequence. So far, there are only few studies showing the presence and effect of cryptic enhancer activities, partly because it requires a detailed dissection of *cis*-regulatory regions to identify cryptic enhancer activities. However given the ability of enhancer activities to change as a result of few mutations and that primed sequences would need fewer mutational events (than non-functional sequences) to alter their enhancer activity, it is possible that cryptic enhancer activities are leading the way to evolutionary trajectories that change enhancer activities rapidly. Hence possible existence of cryptic enhancer activities should be taken into consideration when trying to understand the mechanism of how sometimes few changes can lead to drastic differences in enhancer activities.

How can we elucidate whether enhancers with similar activities are bound by similar sets of transcription factors?

For a full understanding of enhancer structure and activity and how they change over time, it is important to exhaustively identify the transcription factors that, *in vivo*, bind to enhancers with different or similar activities to see how changes in transcription factor binding profiles affect enhancer activity. Subsequently, it is also important to determine the composition of transcription factor binding sites within an enhancer as well as the kinetics of the binding events and how changes in these two affect enhancer activity. So far scientists have collected substantial data on *in vitro* binding of certain transcription factors to particular DNA sequences, however it was found that properties of *in vitro* binding were not necessarily accurate predictors of *in vivo* binding (Wilczynski and Furlong 2010). This is because the cellular environment brings together DNA packaged in chromatin as well as co-factors and other binding partners that can all affect the binding kinetics and specificities of transcription factors. Fortunately, new, more powerful and faster techniques, such as PICh (Proteomics of Isolated Chromatin segments), which can isolate the *in vivo* transcription factor bound DNA regions followed by identification of the bound factors using mass-spectrometry (Déjardin and Kingston 2009), are starting to emerge and will possibly be put in high throughput use in the near future. Overall, in-depth data on the *in vivo* binding specificities of transcription factors can help us understand and perhaps predict the sequence changes that make enhancer activities different or similar. This would involve building more accurate models (than the ones available so far, (Wilczynski and Furlong 2010)), which, based on primary sequence data, can find *cis*-regulatory elements in the genome, predict the expression patterns they would drive as well as how these expression patterns would change as a result of different types of mutations.

Elucidating roles of Y1H-identified transcription factors in regulation of *yellow* expression

Y1H identified, for the first time, a large list of transcription factors binding to *yellow* enhancers from multiple species. This is, however, “only the beginning”. Further functional tests and sequence analyses are necessary to get a full understanding of whether and how a Y1H-identified transcription factor regulates *yellow* expression. Below is a brief description of the appropriate tests and analyses necessary to identify a transcription factor as a direct regulator of and characterizing its role on *yellow* expression.

Does knocking down the transcription factor affect the activity of a particular enhancer sub-element during the developmental stages yellow *is expressed?* Testing effects of a transcription factor on a phenotype (e.g., pigmentation) known to be related to a cognate gene’s function is important for understanding the *in vivo* effects of the transcription factor, at least on the biological pathway the cognate gene is involved in if not directly on the gene itself. However, this type of assay is not sufficient to precisely identify whether and how this transcription alters the gene's expression. One of the best ways of understanding if a transcription factor affects a gene's expression is to knock down the transcription factor in the whole organism, or a particular tissue, and compare the gene's expression in the absence versus presence of the transcription factor. The measurement of expression can be done via a solely quantitative technique (e.g., Reverse Transcription Quantitative Polymerase Chain Reaction), but in order to detect changes in spatial patterns one can use a reporter gene driven by the enhancer of the gene that the transcription factor is thought to control. In the case of *yellow*, one can use the enhancer sub-element-reporter-constructs to document the expression pattern of each enhancer sub-element in the presence versus absence of a particular transcription factor. To be comprehensive, this should be done at various developmental stages important for *yellow* function (e.g., larval and pupal stages).

Does the transcription factor bind to a particular yellow *enhancer sub-element in vitro and/or in vivo?*

Testing whether a transcription factor alters the activity of an enhancer does not show that it directly binds to the enhancer region; it may have indirect effects through regulating other genes upstream of the cognate gene. One can elucidate whether this transcription factor directly binds to the enhancer region using *in vitro* as well as *in vivo* assays. In vitro, Electro Mobility Shift (typically requires that the binding site of the transcription factor being tested is known) or DNase I footprinting (typically requires the transcription factor protein in purified form) assays give a qualitative assessment of direct binding between a protein and DNA molecule. However newer techniques, such as MITOMI (Mechanically Induced Trapping of Molecular Interactions), do not need prior information on the binding site of the transcription factor or the protein itself in purified form, and they can give occupancy data up to 12 bp resolution, which is helpful in identifying the binding site of the transcription factor (Maerkl 2011). MITOMI also gives quantitative binding data by providing the affinity of the transcription factor to different DNA sequences *in vitro*, which is also important for gene expression.

Even though the *in vitro* binding assays are very useful in identifying candidate binding sites and affinity of the transcription factors to certain DNA sequences, not all of these properties prove to be true *in vivo*. Hence, *in vitro* binding assays need to be complemented by *in vivo* assays to fully characterize transcription factor binding. Given the availability of an antibody against a particular transcription factor, Chromatin Immuno Precipitation (ChIP) is very useful in identifying binding and binding sites *in vivo*. However in the absence of proper antibodies, up and coming techniques like PICh can be used to gather in vivo binding information. Once a candidate binding site is determined with *in vitro* techniques, one can also delete or mutate (with nucleotide substitutions) the binding site within the enhancer to see how it affects reporter gene expression when compared to the expression driven by a wild type enhancer.

The above functional tests identifying direct binding and binding sites of a transcription factor within an enhancer, as well as looking at the activity of the enhancer at different developmental stages in the presence versus absence of the transcription factor or its binding site, are the golden standards for determining the role of a transcription factor on an enhancer's activity. Conducting such functional tests for all candidate transcription factors with all *yellow* enhancer sub-elements that they were found to bind to can identify the molecular changes underlying the observed differences in enhancer activities. Overall, understanding how enhancers work and change over evolutionary time requires first identifying enhancers and their corresponding transcription factors, followed by elucidating how changes in enhancer sequences alter transcription factor binding and how this in turn affects enhancer activity.

Conclusion

Given their importance for organismal development, physiology and evolution, I set out to understand how enhancers change over evolutionary time at the molecular and functional level. This is a complex question because there is no known universal code underlying enhancer elements. As a result it is non-trivial to find enhancers in the genome or understand their function because knowledge on such features cannot definitively be achieved in the absence of functional assays. This limits the scope of studies examining enhancers because the necessary functional assays for such studies are typically laborious and time consuming. On the other hand, acquiring more knowledge on enhancer structure and activity and how it changes over evolutionary time necessitates conducting experiments with a broader scope with bigger sample sizes, ideally genome-wide and in multiple species. Only these types of in-depth experiments can help inform *in silico* models to make confident predictions about the complex relationship between the sequence and activity of enhancers.

References

Barolo, S. 2012. Shadow enhancers: Frequently asked questions about distributed cis-regulatory information and enhancer redundancy. BioEssays 34:135-141.

Cande, J., Y. Goltsev, and M. Levine. 2009. Conservation of enhancer location in divergent insects. Proc Natl Acad Sci U S A 106:14414-14419.

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

Déjardin, J., and R. E. Kingston. 2009. Purification of proteins associated with specific genomic Loci. Cell 136:175-186.

Frankel, N., G. K. Davis, D. Vargas, S. Wang, F. Payre, and D. L. Stern. 2010. Phenotypic robustness conferred by apparently redundant transcriptional enhancers. Nature 466:490-493.

Frankel, N., D. F. Erezyilmaz, A. P. McGregor, S. Wang, F. Payre, and D. L. Stern. 2011. Morphological evolution caused by many subtle-effect substitutions in regulatory DNA. Nature 474:598-603.

Maerkl, S. J. 2011. Next generation microfluidic platforms for high-throughput protein biochemistry. Current Opinion in Biotechnology 22:59-65.

Pan, D., S. A. Valentine, and A. J. Courey. 1994. The bipartite D. melanogaster twist promoter is reorganized in D. virilis. Mech Dev 46:41-53.

Pennacchio, L. A., N. Ahituv, A. M. Moses, S. Prabhakar, M. A. Nobrega, M. Shoukry, S. Minovitsky, I. Dubchak, A. Holt, K. D. Lewis, I. Plajzer-Frick, J. Akiyama, S. De Val, V. Afzal, B. L. Black, O. Couronne, M. B. Eisen, A. Visel, and E. M. Rubin. 2006. In vivo enhancer analysis of human conserved non-coding sequences. Nature 444:499-502.

Perry, M. W., A. N. Boettiger, J. P. Bothma, and M. Levine. 2010. Shadow enhancers foster robustness of Drosophila gastrulation. Current Biology 20:1562-1567.

Perry, M. W., A. N. Boettiger, and M. Levine. 2011. Multiple enhancers ensure precision of gap gene-expression patterns in the Drosophila embryo. Proc Natl Acad Sci U S A 108:13570-13575.

Rebeiz, M., N. Jikomes, V. Kassner, and S. Carroll. 2011. Evolutionary origin of a novel gene expression pattern through co-option of the latent activities of existing regulatory sequences. Proc Natl Acad Sci U S A 108:10036-10043.

Schneuwly, S., R. Klemenz, and W. J. Gehring. 1987. Redesigning the body plan of Drosophilaby ectopic expression of the homoeotic gene Antennapedia. Nature 325:816-818.

Sumiyama, K., and N. Saitou. 2011. Loss-of-Function Mutation in a Repressor Module of Human-Specifically Activated Enhancer HACNS1. Molecular Biology and Evolution 28:3005-3007.

Weiszmann, R., A. S. Hammonds, and S. E. Celniker. 2009. Determination of gene expression patterns using high-throughput RNA in situ hybridization to whole-mount Drosophila embryos. Nature Protocols 4:605-618.

Wilczynski, B., and E. E. M. Furlong. 2010. Challenges for modeling global gene regulatory networks during development: Insights from Drosophila. Developmental Biology 340:161-169.