

# **Cis-regulatory evolution integrated the Bric-à-brac transcription factors into a novel fruit fly gene regulatory network**

**Running Header: CRE evolution confers a new function to a pre-existing transcription factor.**

Maxwell J. Roeske<sup>1,#</sup>, Eric M. Camino<sup>1,#</sup>, Sumant Grover<sup>1</sup>, Mark Rebeiz<sup>2</sup>, and Thomas M. Williams<sup>1,3,\*</sup>

<sup>1</sup>Department of Biology, University of Dayton, 300 College Park, Dayton, OH 45469, USA

<sup>2</sup>Department of Biological Sciences, University of Pittsburgh, Pittsburgh, PA 15260

<sup>3</sup>Center for Tissue Regeneration and Engineering at Dayton, University of Dayton, 300 College Park, Dayton, OH 45469, USA

#These authors contributed equally to this work

\*Correspondence: [twilliams2@udayton.edu](mailto:twilliams2@udayton.edu)

## Abstract (150 words)

Gene expression evolution through gene regulatory network (GRN) changes has gained appreciation as a driver of morphological evolution. However, understanding how GRNs evolve is hampered by finding relevant *cis*-regulatory element (CRE) mutations, and interpreting the protein-DNA interactions they alter. We investigated evolutionary changes in the duplicated Bric-à-brac (Bab) transcription factors and a key Bab target gene in a GRN underlying the novel dimorphic pigmentation of *D. melanogaster* and its relatives. It has remained uncertain how Bab was integrated within the pigmentation GRN. Here we show that Bab gained a role in sculpting sex-specific pigmentation through the evolution of binding sites in a CRE of the pigment-promoting *yellow* gene and without any noteworthy changes to Bab protein coding sequences. This work demonstrates how a new trait can evolve by incorporating existing transcription factors into a GRN through CRE evolution, an evolutionary path likely to predominate newly evolved functions of transcription factors.

## Introduction

Transcription factors play central roles in the development and evolution of animal traits by binding to *cis*-regulatory elements to spatially- and temporally-regulate patterns of gene expression (Davidson, 2006; Levine, 2010). Collectively, the complex web of connections between transcription factors and CREs form vast gene regulatory networks (GRN) that govern a tissue's development. As transcription factor genes are generally much older than the traits they impact, a central question of evolutionary developmental biology is the relative role that gene duplication, protein coding and CRE sequence evolution play in the evolution of GRNs for novel traits (Carroll, 2008; Stern and Orgogozo, 2008). Answers to this question require studies of recently evolved traits for which the derived and ancestral states of genes and gene components can be inferred through manipulative studies of GRN function (Rebeiz and Williams, 2011).

One such experimentally tractable trait is the rapidly evolving pigmentation patterns that adorn the abdominal cuticle of *Drosophila melanogaster* and its close relatives (Rebeiz and Williams, 2017). The melanic pigmentation of the dorsal cuticle tergites covering the A5 and A6 abdominal segments of males has been inferred to be a novelty that evolved in the *D. melanogaster* lineage after it diverged from an ancestral monomorphically pigmented lineage within the *Sophophora* subgenus (Jeong et al., 2006). The GRN controlling this trait ultimately instructs the expression of terminal enzyme genes that are required for pigment formation. One enzyme, encoded by the *yellow* gene is activated through a CRE known as the "body element" (Camino et al., 2015; Wittkopp et al., 2002) that possesses at least two binding sites for the Hox factor Abd-B (Jeong et al., 2006). In *D. melanogaster* females, *yellow* expression is repressed

in the A5 and A6 segments through its regulation by the tandem duplicate *bab1* and *bab2* genes (Figure 1A) (Jeong et al., 2006), collectively referred to as Bab. Both paralogs encode proteins that possess a conserved BTB domain that functions in homodimerization and heterodimerization, as well as pipsqueak (psq) and AT-hook domains that together confer *in vitro* DNA-binding capability (Lours et al., 2003). Though the Bab proteins are suspected to function as transcription factors, no direct targets of regulation are known (Jeong et al., 2006).

In the lineage of *D. melanogaster*, Bab expression is suspected to have evolved from a monomorphic pan-abdominal pattern to a sexually dimorphic expression pattern in which both paralogs are female-limited, and absent from males during the latter half of pupal development when enzymes of the pigmentation GRN are deployed (Kopp et al., 2000; Salomone et al., 2013). This dimorphic pattern of Bab expression required changes to two CREs controlling Bab's abdominal epidermis expression (Williams et al., 2008). Though this dimorphic pattern of regulation allows *yellow* to be expressed in the epidermis underlying the pigmented male A5 and A6 segment tergites, several questions remain unanswered pertaining to how and when Bab was incorporated into the sexually dimorphic pigmentation GRN. Specifically, is *yellow* a direct target of Bab repression, and was it a target of regulation prior to the evolution of the dimorphic trait (Gompel and Carroll, 2003)? Moreover, to what extent did gene duplication, protein coding sequence, and CRE evolution contribute to this derived trait?

In this study we sought to characterize the derived functions of Bab1 and Bab2 in the *D. melanogaster* pigmentation GRN and determine the extent to which this derived pigmentation function additionally required evolutionary changes to Bab protein coding

sequences and the 5' *cis*-regulatory region of *yellow* that contains its abdominal CRE. We found that Bab1 and Bab2 bind directly to the *yellow* body CRE to a region required for male-limited enhancer activity. Moreover, the capability of Bab paralogs to function in this repressive manner appears to result not from protein-coding changes, but instead through the evolution of these binding sites which arose contemporaneously with the derived pattern of dimorphic Bab expression. Thus, the origin of the male-specific pigmentation of *D. melanogaster* is an example where evolutionarily conserved transcription factors gained a new function through their integration into a GRN by CRE evolution.

## Results

### Bab1 suppress *yellow* expression through *cis*-regulatory element encodings

We sought to characterize how Bab1 exerts its influence on a minimal 0.6kb body element CRE (yBE0.6) (Camino et al., 2015) (Figure 1B) that drives male-limited GFP reporter transgene expression in the dorsal epidermis of the A5 and A6 abdominal segments (Figure 1C and 1D). This reporter transgene activity matches the spatial, sex-limited, and temporal pattern of abdominal expression of *yellow* (Camino et al., 2015). We designed a set of 10 mutant yBE0.6 CREs (Figure S1) to localize regions responsible for this element's sex-limited activity. In each mutant yBE0.6, we introduced a block of ~70-85 base pairs in which every other base pair possessed non-complementary transversion mutations and compared its activity to the wild type CRE in transgenic *D. melanogaster* (Camino et al., 2015). While 8 of 10 "scanning mutant" CREs showed wild type reporter repression in the female abdomen (Figure S2), we

observed increased expression in the A5 and A6 segments of the SM4 and SM10 mutants (Figure 1E and 1F). Moreover, the increased expression was more pronounced when the SM4 and SM10 mutations were combined in the same reporter (Figure 1G). These results identified two CRE sub-regions that are required to suppress *yellow* expression in the posterior female abdomen, likely through the recruitment of a transcriptional repressor protein.

We speculated that the sequences altered by the SM4 and SM10 mutations normally function to respond either directly or indirectly to the repressive Bab proteins. When Bab1 was ectopically expressed in the dorsal midline of the male abdomen by the GAL4/UAS system (Brand and Perrimon, 1993; Calleja et al., 2000), yBE0.6 reporter expression was largely suppressed (Figure 1H and 1I). However, a yBE0.6 CRE containing the SM4 and SM10 mutations, was unresponsive to ectopic Bab 1 expression (Figure 1J). These data reveal that the SM4 and SM10 CRE regions encode inputs that respond to regulation by Bab proteins.

### **Bab1 directly interacts with multiple *yellow* CRE binding sites**

Bab proteins may suppress the yBE0.6 CRE activity through two major routes: indirect or direct regulation. Bab may indirectly regulate the yBE0.6 CRE by controlling the expression of a transcription factor that interacts with binding sites in the SM4 and SM10 regions. On the other hand, Bab may directly interact with binding sites in the yBE0.6 CRE, acting as a transcriptional repressor. To distinguish between these mechanisms, we performed gel shift assays to see whether Bab specifically interacts with sequences within the SM4 and SM10 regions *in vitro*. Previously it was shown that

the DNA-binding domain (DBD) of Bab1 and Bab2 bound to similar DNA sequences (Lours et al., 2003), therefore we chose to perform our experiments with the Bab1 DBD.

The SM4 region is 70 base pairs (bp) in length (Figure 2A), which we divided into three overlapping sub element regions that were tested for binding by serial two-fold dilutions of a GST fusion protein possessing the Bab1 DNA binding domain (Bab1 DBD, Figure 2B). Of the three regions, only the third probe was substantially bound by the Bab1 DBD (Figure 2C-2E). This binding was DNA-sequence specific, as a scanning mutant probe version failed to similarly shift (Figure 2F). Thus, it seems likely that this 25 base pair segment possesses a site or sites capable of Bab1 DNA-binding.

A previous study showed that the Bab1 DBD preferably bound A/T rich sequences, specifically those with TAA or TA repeats (Lours et al., 2003). Within region 3 are several TA motifs, several of which are part of TAA motifs (Figure 2A). We created a TA>GA mutant probe that removed each TA motif. This probe was not noticeably bound by the Bab1 DBD indicating that some or all of these motifs are necessary features of a Bab1 binding site or sites (Figure 2G). To further localize the sequences necessary for the Bab1 binding, we created four mutant probes within the region (sub1-sub4). We found that three of these mutant probes were still bound and shifted by the Bab1 DBD (Figure 2H-2K). However, the sub2 probe that spanned 9 base-pairs and disrupted two of the TA motifs was not noticeably shifted (Figure 2I). Collectively, this set of gel shift assays supports a direct Bab1 binding mechanism to suppress this CRE sequence (Figure 2A, red sequence; Figure S1).

We performed a similar set of gel shift assays to localize Bab1 binding within the 85 bp SM10 region (Figure 3). We tested three overlapping sub element regions for an

interaction with the Bab1 DBD (Figure 3C-3E), and observed mobility shifts for the first and third regions (Figure 3C and 3E). A scanning mutant version of the third region was bound, albeit to a lesser extent, by Bab1 (Figure 3G), indicating that this *in vitro* binding was not sequence-specific or that the mutant probe created a weak Bab-binding site. Therefore, we focused our attention on the first region for which a scan mutant probe was not noticeably bound by the Bab1 DBD (Figure 3F). We next analyzed four small mutant versions (sub1-sub4) of the first region (Figure 3H-3K). While each mutant probe appeared to be bound to a lesser extent than the wild type probe, the sub3 mutant probe showed little-to-no binding (Figure 3J). These results suggest the presence of at least one additional Bab binding site in the SM10 region. Collectively, our results support a direct model of regulation in which Bab1 functions as a DNA-binding transcription factor that interacts with at least two sites within the yBE0.6 CRE and thereby represses the expression of this CRE in the female abdomen where Bab1 is highly expressed.

## **The biochemical activity of Bab predates its duplication event**

While the *bab* genes have been shown to be sufficient to suppress *D. melanogaster* tergite pigmentation when ectopically expressed (Couderc et al., 2002; Kopp et al., 2000), the individual necessities of *bab1* and *bab2* paralogous genes have not been fully resolved. We created two short hairpin/miRNA (shmiR) transgenic lines that can conditionally and specifically target sequences unique to *bab1* (Table S1) and separately two lines targeting *bab2* (Table S2) for RNA-interference (RNAi). These shmiR transgenes are under the *cis*-regulatory control of Upstream Activating Sequences (UAS), and thus expression can be induced by the GAL4 transcription factor

(Haley et al., 2008). Using a GAL4 insertion into the *pannier* (*pnr*) gene, we drove hairpins specific to a negative control gene (targeting the *mCherry* reporter gene) and to either or both *bab1* and *bab2* along the dorsal midline of the body. Relative to the control (Figure 4A), the *bab1* shmiR transgene containing the siRNA id #3 sequence (Table S1) led to a conspicuous increase in the dorsal medial pigmentation of the female A5 and A6 tergites (Figure 4B), whereas the transgene including the siRNA id #4 sequence (Table S1) resulted in a phenotype not noticeably different from that of the negative control (Figure 4C). For *bab2*, the individual ectopic expression of the shmiR transgenes containing either the siRNA id #12 or #16 sequences (Table S2) resulted in ectopic dorsal medial tergite pigmentation in females (Figure 4D and 4E). Hence, these results demonstrate that suppression of tergite pigmentation in *D. melanogaster* females requires individual contributions from both *bab1* and *bab2*.

The RNAi knockdown of *bab1* and *bab2* each increased pigmentation, suggesting that we would obtain a more expressive phenotype if both paralogs' expression were suppressed. To test this prediction, we created "chained" shmiRs (Haley et al., 2010) to co-express the effective *bab1* shmiR (Id #3) separately with each *bab2* shmiR (Id #12 and #16). We found that these chained shmiR transgenes, when ectopically expressed in the dorsal midline resulted in more expansive ectopic pigmentation phenotypes (Figure 4F and 4G). Notably, for one chained combination, ectopic pigmentation included the A4 tergites of males and females (Figure 4F, red arrowheads). These results show that the individual and combined contributions of *bab1* and *bab2* analog expression are necessary to fully suppress tergite pigmentation.

Not surprisingly, the RNAi phenotypes were most extreme in the female abdomen as the pupal male abdomen lacks significant *bab1* and *bab2* expression (Salomone et al., 2013). Previous studies have shown that both *bab1* and *bab2* are sufficient to suppress male tergite pigmentation when ectopically expressed (Couderc et al., 2002; Kopp et al., 2000). However, direct comparisons of the individual paralogs were hampered by the positional effects associated with random insertion of *bab* paralog open reading frame (ORF) transgenes into different genomic sites. Here, we created transgenes with the *D. melanogaster bab1* and *bab2* ORFs under UAS regulation that were integrated site-specifically into the attP40 site on the 2<sup>nd</sup> chromosome. In the absence of a GAL4 driver, leaky expression of these transgenes resulted in reduced pigmentation of the male A5 and A6 tergites, and a near complete loss of melanic pigmentation on the female A6 tergite (Figure 5A and 5B). When we drove ectopic expression of these ORFs in the dorsal midline by the *pnr-GAL4* chromosome, we failed to recover viable adult males which possessed the UAS-*bab1* ORF transgene, indicating that the ectopic-expression phenotype was lethal. While we obtained fewer than expected offspring with ectopic expression of the *bab2* ORF than expected for independent assortment (Figure S3), some adults were identified. For these specimens, tergite pigmentation was eliminated in the dorsal midline of males, and a split tergite phenotype was seen for both males and females (Figure 5C). These outcomes show that *bab1* and *bab2* have a strong pigment-suppressing capability, though in the genetic background tested, *bab1* ectopic expression was lethal.

The lethality encountered in the ectopic-expression assays likely stems from expressing the *bab* ORFs in the spatial and temporal pattern of the *pnr* gene (Calleja et

al., 2000) which drives strong expression from embryonic stages through pupal development. In order to better visualize the specific effects of *bab* ORF expression on tergite pigmentation, we utilized the *y-GAL4* transgene (Hart, 2013) to drive expression in the abdominal epidermis pattern of the *yellow* gene which begins around 70 hours after puparium formation (hAPF) (Figure S4) based on a 100 hour period of pupal development (Rogers and Williams, 2011). Expression of the *bab1* and *bab2* ORFs driven by the *y-GAL4* chromosome eliminated both lethality and tergite developmental defects, resulting in male adults which entirely lacked melanic pigmentation on the A5 and A6 tergites (Figure 5F and 5G). Thus, by these ectopic expression assays, we find that not only are the Bab paralogs sufficient to suppress pigmentation when ectopically expressed, but that their suppressive capabilities are similar.

### **Bab1 ectopic-expression phenotypes require DNA-binding capability**

The Bab1 and Bab2 proteins possess a conserved domain that includes both pipsqueak (psq) and AT-Hook motifs that function as an *in vitro* DNA-binding domain or DBD (Lours et al., 2003), supporting the notion that these paralogs function as transcription factors *in vivo*. However, bona fide direct targets of either Bab1 or Bab2 have yet to be discovered. Previously, it was shown that the Bab1 DBD failed to bind DNA *in vitro* when possessing non-synonymous mutations in the pipsqueak (psq) motif converting Alanine and Isoleucine amino acids to Glycine and Proline respectively (AI576GP), or when non-synonymous mutations altered a stretch of Arginine, Glycine, and Arginine amino acids in the AT-Hook motif respectively to Aspartic acid, Glycine, and Aspartic acid (RGR627DGD) (Lours et al., 2003). To test whether this compromised

ability to bind DNA *in vitro* has *in vivo* significance, we created a *bab1* ORF transgene that possesses both the psq and AT-Hook mutations (called *bab1* DNA binding mutant or *bab1* DBM, Figure S5) and incorporated this transgene into the same genomic site as our other UAS transgenes. We found that leaky expression of the *bab1* DBM was insufficient to suppress tergite pigmentation (Figure 5D). Moreover, ectopic expression of the Bab1 DBM by the *pnr-GAL4* driver resulted in detectable accumulation of nuclear protein (Figure S6) which did not induce lethality or a pigmentation phenotype (Figure 5E). Similarly, expression of the Bab1 DBM in the *y-GAL4* pattern resulted in males with the wild type melanic tergites (Figure 5H). Collectively, these results lend further support to a model in which DNA binding is required for the *D. melanogaster* Bab paralogs to repress tergite pigmentation.

## **Functional equivalence of Bab paralogs for the suppression of tergite pigmentation**

The currently favored model for the origin of the *D. melanogaster* sexually dimorphic tergite pigmentation posits that it evolved from an ancestor that expressed Bab in a sexually monomorphic manner, and for which melanic pigmentation in males and females was limited (Jeong et al., 2006; Kopp et al., 2000; Salomone et al., 2013). Moreover, CRE evolution has prominently factored into the origin of this dimorphic pigmentation trait, as changes in CREs of *bab* have been previously identified (Williams et al., 2008). The possibility that Bab protein coding sequence evolution has additionally contributed has largely remained untested. To investigate whether the Bab1 and Bab2 proteins have functionally evolved, we created UAS-regulated transgenes possessing

the *D. willistoni bab1* and *D. mojavensis bab2* ORFs (Figure S5). These orthologous protein coding sequences come from fruit fly species presumed to possess the ancestral sexually monomorphic patterns of pigmentation and Bab expression. As seen for the *D. melanogaster* ORF transgenes, leaky expression from the attP2 genomic site of transgene insertion resulted in a similar reduction in male A5 and A6 and female A6 tergite pigmentation (Figure S7A, *D. willistoni bab1*; and S7C, *D. mojavensis bab2*). Pigmentation was dramatically suppressed when these orthologous proteins were ectopically expressed in the dorsal medial midline pattern of *pnr-GAL4*, (Figure S7B, *D. willistoni bab1*; and S7D, *D. mojavensis bab2*) and *y-GAL4* (Figure 5I, *D. willistoni bab1*; and 5J, *D. mojavensis bab2*). These results suggest that the ability of the Bab1 and Bab2 proteins to regulate dimorphic pigmentation did not require noteworthy changes in their protein coding sequences.

The origin of the Bab1 and Bab2 paralogs likely resulted from a duplication event that occurred in the evolutionary history of the Dipteran order (Figure 1A). This scenario is supported by the findings that all species of fruit flies with sequenced genomes and the Tsetse fly *Glossina morsitans* possess *bab1* and *bab2* paralogs, whereas basally branching Dipteran species such as *Anopheles (A.) gambiae* and *Aedes aegypti* only possess a single paralog. Moreover, the genomes of the flour beetle *Tribolium castaneum* and the moth *Bombyx mori*, from orders closely related to Diptera, each possess a single *bab* gene. We were curious as to whether the pigmentation and gene regulatory functions of the derived *bab* paralogs were present in the pre-duplication *bab* gene. To test this hypothesis, we created a UAS-regulated ORF transgene for the *A. gambiae bab* gene to use as a surrogate for the pre-duplication ancestral *bab* gene. We

found that like all other *bab* orthologs tested in this study, leaky expression of the *A. gambiae bab* gene resulted in reduced pigmentation in the male A5 and A6 tergites (Figure S7E). Moreover, forced ectopic expression driven by *pnr-GAL4* and *y-GAL4* both resulted in more extensive repression of tergite pigmentation (Figure S7F and 5K). These outcomes are consistent with the interpretation that the general functional capability of these proteins was ancient and has been generally conserved during the 100 million years or more since the gene duplication event occurred.

We sought to determine whether the functional equivalence of the distantly related Bab orthologs can also be seen at the level of target gene regulation. Here, we focused on the capability of the orthologs to repress the expression of the Enhanced Green Fluorescent Protein (EGFP) reporter transgene expression driven by the sequence 5' of *yellow* exon 1 that contains the wing element and body element CREs (Figure 6A) from the dimorphic species *D. melanogaster* (Figure 6B-6G) and *D. malerkotliana* (Figure 6H-6M). Orthologs were ectopically expressed in the pupal abdomens of *D. melanogaster* under the control of the *y-GAL4* transgene and the intensity of EGFP expression in the A5 and A6 segment epidermis relative to that for specimens ectopically expressing the Bab1 DBM was measured (Figure 6B and 6H). We observed the *Drosophila* and *Anopheles* orthologs were similarly capable of repressing reporter transgenes regulated by the *yellow* regulatory regions of both dimorphic species. Thus, it can be concluded that the functional equivalency of distantly related Bab orthologs extends to the ability to regulate a derived target gene in a *Drosophila* lineage.

## **The gain of direct Bab regulation required CRE evolution for *yellow***

The male-specific pigmentation of the A5 and A6 tergites is thought to be a derived state in the lineage of *D. melanogaster* after it diverged from monomorphic lineages such as that of the *willistoni* and possibly the *obscura* species groups (Figure 8B). This may have occurred by two distinct possible routes. First, the regulation of *yellow* expression by Bab might predate the dimorphic pattern of tergite pigmentation, and thus, when Bab expression evolved dimorphism, the *yellow* gene became restricted to males. Alternately, the regulation of *yellow* by Bab may have originated contemporaneously with the evolution of dimorphic Bab expression. We compared the *D. melanogaster* Bab-binding sites from the SM4 and SM10 regions to the orthologous gene regions from species descending from either a dimorphic or monomorphic pigmented ancestor (Figure 8A). This analysis revealed a comparable degree of sequence conservation to the derived binding sites for the Hox transcription factor Abd-B that were previously shown to be a key event in the evolution of male-limited *yellow* expression (Jeong et al., 2006). However, there is abundant sequence divergence among species that descend from the dimorphic lineage (Figure 8A, node 1), and a near complete absence of sequence conservation with species possessing the ancestral monomorphic trait. Thus, it is possible that an ancestral regulatory linkage between Bab and *yellow* is obscured by the turnover and displacement (Hare et al., 2008; Ludwig et al., 2000; Swanson et al., 2011) of Bab-binding sites to other regions of the body element.

To infer the antiquity of Bab-repression at *yellow*, we compared the capabilities of the *D. melanogaster* Bab orthologs to affect the CRE activities of *yellow* 5' regulatory

regions (containing both the wing and body elements) from dimorphic species (*D. melanogaster* and *D. malerkotliana*) and ancestrally monomorphic species (*D. pseudoobscura* and *D. willistoni*) (Figure 7). For regulatory sequences derived from dimorphic species, reporter expression in A5 and A6 segments was strikingly reduced in the presence of either ectopic Bab1 or Bab2 compared to the Bab1 DBM control (Figure 7A-A'' and 7B-B''). In contrast, the regulatory sequences from monomorphic species showed modest and no apparent reduction of *D. pseudoobscura* and *D. willistoni* CREs, respectively. Two additional observations can be made. One being the even greater level of regulatory-activity repression for the *D. malerkotliana yellow* regulatory region (33%±1% in the presence of ectopic Bab1 or Bab2) than for *D. melanogaster* (56%±1% and 53%±1% respectively in the presence of ectopic Bab1 and Bab2). Second, the regulatory activity of the *D. pseudoobscura* was modestly repressed in the presence of Bab (85%±2% and 71%±2% respectively in the presence of ectopic Bab1 and Bab2).

While A5 and A6 *yellow* expression is largely governed by the body element in males, the wing element does direct a more moderate level of expression throughout the abdomen and in posterior stripes for each segment. To see whether the wing element might also respond to Bab expression and thus harbor Bab-binding sites, we compared the levels of expression for the reporter transgenes in the A3 segment in which expression is driven exclusively by the wing element (Figure S8). While the *D. melanogaster* wing element showed a slight reduction in activity when in the presence of ectopic Bab1 or Bab2 compared to the Bab1 DBM control, no apparent reduction in activity was observed for the transgenes with either the *D. malerkotliana*, *D. pseudoobscura*, or *D. willistoni yellow* regulatory sequences (Figure S8). Collectively,

these results suggest that the direct Bab regulation of *yellow* evolved specifically within the dimorphic body element, coincident with the evolution of the dimorphic pigmentation trait.

## Discussion

Here, we investigated the functions and evolution of the paralogous *D. melanogaster* Bab1 and Bab2 proteins that perform a key regulatory role in a GRN controlling a male-specific pigmentation trait. Though these two paralogs descend from an ancient duplication event (Figure 1A), our results show that their ability to function in *D. melanogaster* pigmentation required little-to-no alteration in the functional capability of the Bab proteins. Rather, our data point to the evolution of binding sites in a CRE of a key pigmentation gene *yellow* to which these proteins bind through their DNA binding domains (Figure 8A). These conclusions are supported by multiple lines of evidence: (1) identification of regions in the *yellow* body CRE that mediate sexual dimorphism, (2) *in vitro* binding of Bab proteins to these sequence, (3) the ability of pre-duplicate Bab proteins to suppress the *yellow* body CRE of dimorphic species, and (4) the relative inability of these proteins to exert similar effects on CREs of species whose lineages predate the evolution of this trait. These findings represent the first-documented direct target sites of the Bab proteins, sites which arose coincident with the evolution of sexually dimorphic Bab expression patterns in the abdomen (Figure 8B). Thus, our results provide a clear example in which multiple tiers of a complex GRN evolved to produce a Hox-regulated trait while preserving the genetic toolkit of regulatory and differentiation genes.

## The evolution of the Bab paralogs

The *D. melanogaster bab* locus provides an interesting example in which the protein-coding and regulatory divergence of duplicated genes can be compared. The phylogenetic distribution of *bab* paralogs supports an estimated timing of the duplication event to around 125 million years since the common ancestor of the fruit fly, Tsetse fly, and Hessian fly split from the lineage of the sandfly and mosquitoes (Figure 1A) (Wiegmann et al., 2011). Duplicate genes may sub-functionalize, neo-functionalize, or be lost through pseudogenization (Lynch and Conery, 2000). Since this duplication event, both *bab* paralogs have been maintained in the genomes of distantly related fruit fly species (Clark et al., 2007; Richards et al., 2005), and in species from related families whose genomes have been sequenced (Giraldo-Calderon et al., 2015; Kriventseva et al., 2015). Here, we showed that both *D. melanogaster bab* paralogs and orthologs from other fruit fly and a mosquito species can similarly impact the development of melanic tergite pigmentation when ectopically expressed in *D. melanogaster* (Figure 5), and each gene can induce a similar split tergite phenotype (Figure 5 and S7). These outcomes suggest for at least these two phenotypes that their protein coding regions are functionally equivalent.

In contrast to the conserved protein functionality that we observed for the Bab paralogs, some divergent patterns of expression have been found for the paralogs in *D. melanogaster*, consistent with a role for neo- or sub-functionalization (Couderc et al., 2002). Yet, the majority of *bab* paralog expression patterns appear to be common to both paralogs (Couderc et al., 2002; Rogers et al., 2013; Salomone et al., 2013), including the pupal abdominal epidermis, for which expression is governed by two

shared CREs (Williams et al., 2008). It has been found that heterozygous *bab* null females have a more male-like pattern of tergite pigmentation compared to wild type females, and the homozygous null pigmentation phenotype is more or less equivalent to that of males. (Rogers et al., 2013). Here we showed that the RNA-i reduction of expression for either *bab1* or *bab2* resulted in more male-like pigmentation pattern on the female abdomen, and RNA-i for both paralogs simultaneously resulted in a more pronounced male-like phenotype (Figure 4). While qualitative differences through CRE functional divergence must have occurred to drive the divergent paralog expression in some tissues, the need for a higher overall quantity of expressed Bab protein seems to be key for Bab's role in the GRN generating the derived dimorphic pattern of abdomen pigmentation.

## **Bab and its history in a pigmentation gene regulatory network**

The stark dimorphism between the melanic pigmentation of *D. melanogaster* male and female abdominal tergites represents a trait whose origin has now been resolved to the level of its GRN connections. Dimorphic pigmentation is thought to have derived from a monomorphic ancestral state in the lineage of *D. melanogaster* after it diverged from that for *D. willistoni* and perhaps even as recently as the *D. pseudoobscura* split in the *Sophophora* subgenus (Jeong et al., 2006; Salomone et al., 2013). Assuming this scenario is generally correct, then when and how did Bab become a part of this GRN? One plausible explanation is that Bab regulated the expression of pigmentation genes prior to the emergence of this dimorphic trait. Perhaps as a part of an antecedent dimorphic GRN. One study provided data consistent with this scenario,

showing that Bab2 expression often, but not always, displayed an anti-correlation to where melanic pigmentation developed on fruit fly tergites, including non-*Sophophora* species (Gompel and Carroll, 2003). If Bab had an ancient role in regulating the expression of pigmentation genes such as *yellow*, then dimorphic pigmentation could have evolved by re-deploying a conserved Bab-responsive CRE in the abdomen.

However, various data are difficult to reconcile with this re-deployment model. First, the melanic species *D. virilis* expresses Bab1 and Bab2 throughout the male and female abdomens suggesting an inability of Bab to suppress pigmentation genes such as *yellow* in this species (Salomone et al., 2013). Furthermore, the regulation of *yellow* by Bab in *D. melanogaster* is limited to the body element CRE in which we found Bab binding sites. This is significant as clearly identifiable orthologs to the *D. melanogaster* body element remains difficult to identify in more distantly related fruit fly species hailing from monomorphically pigmented lineages (Figure 8A), though the wing element and its activity has remained comparatively well-conserved. Thus, there is no evidence for an ancient linkage between Bab and an antecedent of the *yellow* body element. While an ancestral *bab-yellow* linkage may exist for a yet unidentified trait, the activation of *yellow* in the well-studied male-specific spot that adorns the wings of *D. biarmipes* is *bab*-independent (Arnoult et al., 2013; Gompel et al., 2005). Finally, the *D. melanogaster* body element possess derived binding sites for both Bab and the Hox factor Abd-B (Figure 8A) (Jeong et al., 2006). Thus, the co-option of either Bab or Abd-B would be insufficient to account for the origin of this dimorphic trait.

Based upon the available data, it seems much more likely that Bab was integrated into an antecedent GRN to play a key role in differentiating the expression

outcomes between males and females. This integration involved the remodeling of existing CREs controlling Bab expression in the abdominal epidermis (Rogers et al., 2013; Williams et al., 2008), and the acquisition of *yellow* as a direct target of regulation through the formation of binding sites in the emergent body element (Figure 8). At this point it is unclear whether dimorphic Bab expression preceded the evolution of Bab sites in the *yellow* body element CRE or if the Bab sites evolved first. One hint to this puzzle may lie in *D. pseudoobscura*, a species whose *yellow* body element CRE is mildly Bab-responsive (Figure 7 and S8), but for which Bab expression retains its ancestral monomorphic expression. This might suggest that the capability to respond (albeit weakly) to Bab evolved first, followed by the evolution of dimorphic Bab expression patterns. Further, the evolutionary connections identified here represent a subset of the connections downstream of Bab. While the *yellow* body CRE contains separable activating and repressing inputs, the gene *tan*, which is co-expressed with *yellow* appears to have a very different encoding for dimorphism (Camino et al., 2015). Extensive mutagenesis of the *tan* MSE failed to find any mutations that relaxed dimorphic expression, suggesting that activating and repressing inputs are overlapping, or closely situated in this CREs DNA sequence. Future studies of this GRN will illuminate how the network downstream of Bab was elaborated.

## **The incorporation of old transcription factors into new networks**

We suggest that the increased complexity of the dimorphic pigmentation GRN through the integration of the Bab transcription factors by CRE evolution will exemplify a common mechanism whereby increasingly sophisticated GRNs have come about to

regulate traits throughout the animal kingdom. The vast majority of transcription factor binding specificities remain conserved over long evolutionary periods (Nitta et al., 2015), and many of these factors are functionally equivalent between distantly related taxa. The exceptions to this trend may represent rare examples of transcription factor diversification that occurred in the distant past, and thus are limited to a vanishingly small number of traits, or may represent examples of developmental systems drift in which the molecular mechanisms change, but the outcome remains the same. Studying more recent trait divergence allows one to more clearly discern phenotypically relevant evolutionary changes from those involving systems drift. Tests of regulatory sequence divergence are particularly hampered by drift, as sequence divergence is rapid and CREs from distantly related taxa often work poorly in heterologous transgenic environments. Thus, further comparisons of genetically tractable traits that arose over similarly recent timescales, in which protein coding and *cis*-regulatory divergence can be directly compared *in vivo* are required to unveil the nature of this broader trend.

## Materials and Methods

### Fly Stocks and Genetic Crosses

All fly stocks were cultured at 22°C using a sugar food medium (Salomone et al., 2013). The yBE0.6, yBE0.6 SM4, yBE0.6 SM10, and SM4+10 reporter transgenes utilized in GAL4/UAS experiments were each inserted into the attP40 site (Camino et al., 2015; Markstein et al., 2008). GAL4 expression was driven in the pattern of the *pannier* (*pnr*) gene using the *pnr*-GAL4 chromosome (Calleja et al., 2000) and the pupal abdominal epidermis pattern (Jeong et al., 2006; Wittkopp et al., 2002) of the *yellow*

gene using the  $\gamma$ -GAL4 transgene (Hart, 2013). The *pnr-GAL4* (BDSC ID#3039) and  $\gamma$ -GAL4 (BDSC ID#44267) fly stocks were obtained from the Bloomington Drosophila Stock Center. A UAS-mCherry dsRNA line (BDSC ID#35785) was used as a negative control in the RNA-interference experiments. The reporter transgenes containing orthologous sequences 5' of the *yellow* first exon adjacent to a minimal *hsp70* promoter and the coding sequence of the EGFP-NLS reporter protein were integrated into the attP2 site on the *D. melanogaster* 3<sup>rd</sup> chromosome and whose construction was previously described (Camino et al., 2015; Groth et al., 2004).

## **Recombinant Protein Production and Gel Shift Assays**

The protein coding sequence for amino acids 490-672 of the *D. melanogaster* Bab1 protein was cloned into the *Bam*HI and *Not*I sites of the pGEX4T-1 vector in order to express an N-terminal GST-fusion protein that has the AT-Hook and psq domains and that possesses DNA binding capability (Lours et al., 2003). This vector was transformed into the BL21 DE3 *E. coli* strain (New England Biolabs) and recombinant protein was purified by a standard protocol (Williams et al., 1995) with slight modifications. In brief, an overnight bacterial culture was grown at 37°C in LB media with 200 µg/ml Ampicillin. This culture was added to 225 ml of a rich LB media (2% Tryptone, 1% Yeast Extract, and 1% sodium chloride) and grown at 37°C. After 1 hour of growth, protein expression was induced by adding IPTG to a final concentration of 0.5mM, and cultured for an additional 3 hours. Bacteria were then pelleted by centrifugation, media decanted, and bacterial pellets frozen at -74°C. Bacteria pellets were thawed on ice and resuspended in ice cold STE buffer containing protease

inhibitors (Thermo Scientific). After a 15 minute incubation on ice, DTT was added to 5 mM and Sarkosyl to 1.5%. The bacterial suspension was subjected to 4 rounds of sonication on ice at 33 amps for 1 minute each round with, and a 1 minute rest between rounds. 1 ml of glutathione agarose (Thermo Scientific) was then added to the suspension and allowed to mix for 15 minutes with nutation at 4°C. The glutathione-agarose was then washed 7 times with ice cold PBS. GST-Bab1 DNA-binding domain (DBD) fusion protein was eluted from the glutathione-agarose by seven 1.5 ml aliquots of protein elution buffer (75 mM Hepes pH 7.4, 150 mM NaCl, 20 mM reduced glutathione, 5 mM DTT, and 0.1% Triton X-100). Collected aliquots were combined and concentrated using Vivaspin 20 spin columns with a 100,000 MWCO (Sartorius). The purified GST-Bab1 DBD protein was snap frozen using a dry ice ethanol bath, and stored in aliquots at -74°C.

Reverse complementary oligonucleotides were synthesized (Integrated DNA Technologies) that contain wild type or mutant yBE0.6 sequences (Tables S3 and S4). Gel shift assays were done as previously described (Camino et al., 2015; Rogers et al., 2013). In brief, all oligonucleotides were biotin-labeled on their 3' end using the DNA 3' End Biotinylation Kit (Thermo Scientific) using the manufacture's protocol. Biotin-labeled complementary oligonucleotides were annealed to form double stranded probes, and labeling efficiency was evaluated by the manufacturer's quantitative Dot Blot assay. Binding reactions included 25 fmol of probe and GST-Bab1 DBD protein in General Footprint Buffer (working concentration of 50 mM HEPES pH 7.9, 100 mM KCl, 1 mM DTT, 12.5 mM MgCl<sub>2</sub>, 0.05 mM EDTA, and 17% glycerol). For each probe, separate binding reactions were done that included 4,000 ng, 2,000 ng, 1,000 ng, 500 ng, and 0

ng of the GST-Bab1 DBD protein. Binding reactions were carried out for 30 minutes at room temperature and then size separated by electrophoresis through a 5% non-denaturing polyacrylamide gel for 60 minutes at 200 V. Following electrophoresis, the samples were transferred and cross linked to a Hybond-N+ membrane (GE Healthcare Amersham) for chemiluminescent detection using the Chemiluminescent Nucleic Acid Detection Module and manufacture's protocol (Thermo Scientific). Chemiluminescent images were recorded using a BioChemi gel documentation system (UVP).

### ***bab1* and *bab2* shmiR expressing transgenes**

The open reading frame (ORFs) for *D. melanogaster bab1* and *bab2* were obtained from NCBI accession numbers NM\_206229 and NM\_079155.3 respectively. From these, nucleotide guide sequences were designed using the Designer of Small Interfering RNA (DSIR) algorithm (Vert et al., 2006) that is accessible at: <http://biodev.extra.cea.fr/DSIR/DSIR.html>. For *bab1*, the eleven rows of output were included in Table S1. For *bab2*, there were 54 rows of output, sorted by descending Corrected Score, and the top 19 rows of output presented in Table S2. To make sure shmiRs lack the same seed residues (nucleotides 2-8) as those present in known miRNAs, we searched candidate guide sequences against a miRNA database (<http://mirbase.org/search.shtml>). Search sequences were set to "Mature miRNAs, E-value cutoff of "10", Maximum hits of 100, and results were shown for "fly".

Previously it was shown that a shmiR can induce phenotypes in transgenic flies when the guide shares at least 16-21 base pairs of contiguous sequence to the target gene (Haley et al., 2010). Thus we sought the highest scoring "Guide" sequences for

which for which fewer than 16 contiguous bases match a heterologous exon sequence in the *D. melanogaster* genome. Guide sequences were evaluated in a BLAST search of the *D. melanogaster* genome (<http://flybase.org/blast/>) with the word size set to 7. The genomic position of the BLAST hits were identified using the GBrowse feature. An RNAi transgene targeting *bab1* was created by the Transgenic RNAi Project (TRiP) at Harvard Medical School that included the sequence identified here as siRNA\_id 1 (Table S1) that we have found to be ineffective at suppressing *bab1* expression. Thus, this guide sequence was excluded from further consideration here. For *bab1* we elected to create shmiRs with the siRNA 3 and siRNA 4 sequences which each have a 21 base pair match to a sequence in the *bab1* 1<sup>st</sup> exon. For *bab2* we elected to create shmiRs with the siRNA 16 and siRNA 12 sequences, each which have a 21 base sequence that matches a sequence in the 2<sup>nd</sup> exon of *bab2*. The *bab1* and *bab2* shmiRs were designed to possess two essential mismatches to maintain a miR-1 stem-loop structure (Haley et al., 2010), and oligonucleotides were designed for annealing that have *NheI* and *EcoRI* overhangs for cloning into the pattB-NE3 vector (Table S5). The annealed oligonucleotides were cloned into the *NheI* and *EcoRI* sites of the pattB-NE3 vector, and successful cloning was verified by Sanger sequencing using the pUASTR1 primer (5' CCCATTCATCAGTTCCATAGGTTG 3'). pattB-NE3 vectors containing an shmiR guide sequence were site-specifically integrated into the *D. melanogaster* attP2 landing site (Groth et al., 2004) by standard protocol (Best Gene Inc.).

## Chaining shmiRs to target *bab1* and *bab2*

shmiR chains were created in two steps. First the bab1\_3 shmiR was removed from the pattB-NE3 vector by HindIII and BamHI digestion and the excised piece was subcloned into the pHB vector (Haley et al., 2010, 2008). The shmiR piece was then amplified from the pHB vector using the M13F and M13R primers. This PCR product was digested with KpnI and SpeI restriction endonucleases and then cloned into the *KpnI* and *XbaI* sites of the pattB-NE3 vectors containing the bab2 siRNA 16 and the vector containing the bab2 siRNA 12 sequence. For each vector, the presence of the tandem shmiR sequences was verified by Sanger sequencing in separate reactions with the PUASTR1 (5' CCCATTCATCAGTTCCATAGGTTG 3') and PUASTF1 (5' ACCAGCAACCAAGTAAATCAACTG3') primers. These chained shmiR transgenes were injection into *D. melanogaster* embryos for site-specific integration into the attP2 site on the 3<sup>rd</sup> chromosome to make transgenic stocks (Groth et al., 2004).

### ***bab* open reading frame transgenes**

The ORFs for *D. melanogaster bab1* and *bab2*, *D. willistoni bab1* (GK16863-PA), *D. mojavensis bab2* (GI12710), and *Anopheles (A.) gambiae bab* (AGAP006018-RA) were customized for gene synthesis by GenScript Incorporated. We added a Syn21 translational enhancer (Pfeiffer et al., 2012) 5' of each ORF's initiator ATG, and an additional nonsense codon was added just 3' of the endogenous one. The ORFs were flanked by a 5' *EcoRI* site and a 3' *NotI* site. The *A. gambiae* ORF had the coding sequence for the V5 epitope tag added after the initiator codon. The DNA sequences were modified with synonymous substitutions as needed in order to optimize for gene synthesis. The synthesized sequences can be found in Figure S5. After synthesis, the

ORFs were removed from the pUC57 vector backbone and subcloned into the *EcoRI* and *NotI* sites of a pUAST vector modified to possess an attB site (called pUMA) for site-specific transgene integration. All ORFs are under the regulatory control of the vector's upstream activating sequences (UAS sites) to allow for conditional ORF expression by the GAL4/UAS system (Brand and Perrimon, 1993).

A *bab1* DNA-binding mutant (DBM) ORF was created that possesses non-synonymous mutations in the Pipsqueak and AT-Hook motifs that results in a protein lacking its *in vitro* DNA-binding capability (Lours et al., 2003). The coding sequence for the *bab1* Pipsqueak and AT-Hook motifs are flanked by *AscI* and *BamHI* restriction endonuclease sites. We designed a coding sequence within this sequence that includes mutations altering codons with these two protein domains (Figure S5). After synthesis, this mutant sequence was removed from the pUC57 vector by *BamHI* and *AscI* digestion. The liberated fragment was swapped into the place of the wild type sequence in the pUC57 *bab1* ORF vector. The full length *bab1* DBM ORF was then removed from the pUC57 vector by *EcoRI* and *NotI* digestion and subcloned into the pUMA vector. All ORF transgenes were site-specifically integrated into the *D. melanogaster* attP40 site (Markstein et al., 2008) on the 2<sup>nd</sup> chromosome.

## Imaging and analysis of abdomens

Images of the adult *D. melanogaster* abdomen pigmentation patterns were taken with an Olympus SZX16 Zoom Stereoscope equipped with an Olympus DP72 digital camera. Specimens were prepared from 2-5 day old adults. The expression patterns of EGFP-NLS reporter transgenes were recorded as projection images by an Olympus

Fluoview FV 1000 confocal microscope and software. The regulatory activities of the *yellow* 5' sequences were imaged at ~85 hAPF, a time point when endogenous *yellow* expression occurs in the abdominal epidermis (Camino et al., 2015; Jeong et al., 2006). Representative images were selected from the replicate specimens for display in figures presenting reporter transgene expressions. All images underwent the same modifications as the control specimens using Photoshop CS3 (Adobe).

Quantitative comparisons of the levels of EGFP-NLS reporter expression were performed similar to that previously described for another CRE (Camino et al., 2015; Rogers et al., 2013; Rogers and Williams, 2011; Williams et al., 2008). In brief, for each reporter transgene in a genetic background ectopically expressing a Bab protein, EGFP expression was imaged from five independent replicate specimens at developmental stage of ~85 hours after puparium formation (hAPF) for *D. melanogaster* grown at 25°C. Imaging was done with confocal microscope settings optimized so that few pixels were saturated when EGFP-NLS expression was driven in a control background ectopically expressing the DNA-binding mutant form of Bab1 (Bab1 DBM) under the regulation of the *y-GAL4* chromosome. For each confocal image, a pixel value statistic was determined for the dorsal epidermis of the A4 and A5 segments and separately the A3 segment using the Image J program (Abràmoff et al., 2004). In Figure 7, the mean pixel values were converted into a % regulatory activity with its standard error of the mean (SEM). The activities were normalized to the activity in the background expressing the Bab1-DBM which were considered to be 100%. All pixel values of replicate specimens are presented in Figure S8.

# Acknowledgments

This study benefited from stocks obtained from the Bloomington Drosophila Stock Center (NIH P40OD018537), the San Diego Drosophila Stock Center, and from Sean B. Carroll. We are grateful for Michael Levine sharing the pattB-NE3 and pHB vectors to make shmiR transgenes. MJR was supported in part by the Berry Summer Thesis Institute and Honors Program at the University of Dayton. EMC and SG were supported by fellowships from the University of Dayton Graduate School. TMW and MR were supported by a grant from the National Science Foundation (IOS-1555906). MR's work on Drosophila pigmentation was additionally supported by a grant from the National Institutes of Health (5R01GM114093-02). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

# References

- Abràmoff, M.D., Hospitals, I., Magalhães, P.J., Abràmoff, M., 2004. Image Processing with ImageJ. Biophotonics Int. 11, 36–42.
- Arnoult, L., Su, K., Manoel, D., Minervino, C., Magrina, J., Gompel, N., Prud'homme, B., 2013. Emergence and Diversification of Fly Pigmentation Through Evolution of a Gene Regulatory Module. Science 339, 1423–1426. doi:10.1126/science.1233749
- Brand, A.H., Perrimon, N., 1993. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. Development 118, 401–15.
- Calleja, M., Herranz, H., Estella, C., Casal, J., Lawrence, P., Simpson, P., Morata, G., 2000. Generation of medial and lateral dorsal body domains by the pannier gene of

*Drosophila*. Development 127, 3971–80.

Camino, E.M., Butts, J.C., Ordway, A., Vellky, J.E., Rebeiz, M., Williams, T.M., 2015.

The Evolutionary Origination and Diversification of a Dimorphic Gene Regulatory Network through Parallel Innovations in cis and trans. PLOS Genet. 11, e1005136. doi:10.1371/journal.pgen.1005136

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

Clark, A.G., Eisen, M.B., Smith, D.R., Bergman, C.M., Oliver, B., Markow, T.A., Kaufman, T.C., Kellis, M., Gelbart, W., Iyer, V.N., Pollard, D. a, Sackton, T.B., Larracuenta, A.M., Singh, N.D., Abad, J.P., Abt, D.N., Adryan, B., Aguade, M., Akashi, H., Anderson, W.W., Aquadro, C.F., Ardell, D.H., Arguello, R., Artieri, C.G., Barbash, D. a, Barker, D., Barsanti, P., Batterham, P., Batzoglou, S., Begun, D., Bhutkar, A., Blanco, E., Bosak, S. a, Bradley, R.K., Brand, A.D., Brent, M.R., Brooks, A.N., Brown, R.H., Butlin, R.K., Caggese, C., Calvi, B.R., Bernardo de Carvalho, A., Caspi, A., Castrezana, S., Celniker, S.E., Chang, J.L., Chapple, C., Chatterji, S., Chinwalla, A., Civetta, A., Clifton, S.W., Comeron, J.M., Costello, J.C., Coyne, J. a, Daub, J., David, R.G., Delcher, A.L., Delehaunty, K., Do, C.B., Ebling, H., Edwards, K., Eickbush, T., Evans, J.D., Filipski, A., Findeiss, S., Freyhult, E., Fulton, L., Fulton, R., Garcia, A.C.L., Gardiner, A., Garfield, D. a, Garvin, B.E., Gibson, G., Gilbert, D., Gnerre, S., Godfrey, J., Good, R., Gotea, V., Gravely, B., Greenberg, A.J., Griffiths-Jones, S., Gross, S., Guigo, R., Gustafson, E. a, Haerty, W., Hahn, M.W., Halligan, D.L., Halpern, A.L., Halter, G.M., Han, M. V, Heger, A., Hillier, L., Hinrichs, A.S., Holmes, I., Hoskins, R. a, Hubisz, M.J., Hultmark, D.,

713 Huntley, M. a, Jaffe, D.B., Jagadeeshan, S., Jeck, W.R., Johnson, J., Jones, C.D.,  
714 Jordan, W.C., Karpen, G.H., Kataoka, E., Keightley, P.D., Kheradpour, P.,  
715 Kirkness, E.F., Koerich, L.B., Kristiansen, K., Kudrna, D., Kulathinal, R.J., Kumar,  
716 S., Kwok, R., Lander, E., Langley, C.H., Lapoint, R., Lazzaro, B.P., Lee, S.-J.,  
717 Levesque, L., Li, R., Lin, C.-F., Lin, M.F., Lindblad-Toh, K., Llopart, A., Long, M.,  
718 Low, L., Lozovsky, E., Lu, J., Luo, M., Machado, C. a, Makalowski, W., Marzo, M.,  
719 Matsuda, M., Matzkin, L., McAllister, B., McBride, C.S., McKernan, B., McKernan,  
720 K., Mendez-Lago, M., Minx, P., Mollenhauer, M.U., Montooth, K., Mount, S.M., Mu,  
721 X., Myers, E., Negre, B., Newfeld, S., Nielsen, R., Noor, M. a F., O’Grady, P.,  
722 Pachter, L., Papaceit, M., Parisi, M.J., Parisi, M., Parts, L., Pedersen, J.S., Pesole,  
723 G., Phillippy, A.M., Ponting, C.P., Pop, M., Porcelli, D., Powell, J.R., Prohaska, S.,  
724 Pruitt, K., Puig, M., Quesneville, H., Ram, K.R., Rand, D., Rasmussen, M.D., Reed,  
725 L.K., Reenan, R., Reily, A., Remington, K. a, Rieger, T.T., Ritchie, M.G., Robin, C.,  
726 Rogers, Y.-H., Rohde, C., Rozas, J., Rubenfield, M.J., Ruiz, A., Russo, S.,  
727 Salzberg, S.L., Sanchez-Gracia, A., Saranga, D.J., Sato, H., Schaeffer, S.W.,  
728 Schatz, M.C., Schlenke, T., Schwartz, R., Segarra, C., Singh, R.S., Sirot, L., Sirota,  
729 M., Sisneros, N.B., Smith, C.D., Smith, T.F., Spieth, J., Stage, D.E., Stark, A.,  
730 Stephan, W., Strausberg, R.L., Strepel, S., Sturgill, D., Sutton, G., Sutton, G.G.,  
731 Tao, W., Teichmann, S., Tobari, Y.N., Tomimura, Y., Tsolas, J.M., Valente, V.L.S.,  
732 Venter, E., Venter, J.C., Vicario, S., Vieira, F.G., Vilella, A.J., Villasante, A.,  
733 Walenz, B., Wang, J., Wasserman, M., Watts, T., Wilson, D., Wilson, R.K., Wing,  
734 R. a, Wolfner, M.F., Wong, A., Wong, G.K.-S., Wu, C.-I., Wu, G., Yamamoto, D.,  
735 Yang, H.-P., Yang, S.-P., Yorke, J. a, Yoshida, K., Zdobnov, E., Zhang, P., Zhang,

736 Y., Zimin, A. V, Baldwin, J., Abdouelleil, A., Abdulkadir, J., Abebe, A., Abera, B.,  
737 Abreu, J., Acer, S.C., Aftuck, L., Alexander, A., An, P., Anderson, E., Anderson, S.,  
738 Arachi, H., Azer, M., Bachantsang, P., Barry, A., Bayul, T., Berlin, A., Bessette, D.,  
739 Bloom, T., Blye, J., Boguslavskiy, L., Bonnet, C., Boukhgalter, B., Bourzgui, I.,  
740 Brown, A., Cahill, P., Channer, S., Cheshatsang, Y., Chuda, L., Citroen, M.,  
741 Collymore, A., Cooke, P., Costello, M., D'Aco, K., Daza, R., De Haan, G., DeGray,  
742 S., DeMaso, C., Dhargay, N., Dooley, K., Dooley, E., Doricent, M., Dorje, P.,  
743 Dorjee, K., Dupes, A., Elong, R., Falk, J., Farina, A., Faro, S., Ferguson, D., Fisher,  
744 S., Foley, C.D., Franke, A., Friedrich, D., Gadbois, L., Gearin, G., Gearin, C.R.,  
745 Giannoukos, G., Goode, T., Graham, J., Grandbois, E., Grewal, S., Gyaltsen, K.,  
746 Hafez, N., Hagos, B., Hall, J., Henson, C., Hollinger, A., Honan, T., Huard, M.D.,  
747 Hughes, L., Hurhula, B., Husby, M.E., Kamat, A., Kanga, B., Kashin, S.,  
748 Khazanovich, D., Kisner, P., Lance, K., Lara, M., Lee, W., Lennon, N., Letendre, F.,  
749 LeVine, R., Lipovsky, A., Liu, X., Liu, J., Liu, S., Lokyitsang, T., Lokyitsang, Y.,  
750 Lubonja, R., Lui, A., MacDonald, P., Magnisalis, V., Maru, K., Matthews, C.,  
751 McCusker, W., McDonough, S., Mehta, T., Meldrim, J., Meneus, L., Mihai, O.,  
752 Mihalev, A., Mihova, T., Mittelman, R., Mlenga, V., Montmayeur, A., Mulrain, L.,  
753 Navidi, A., Naylor, J., Negash, T., Nguyen, T., Nguyen, N., Nicol, R., Norbu, C.,  
754 Norbu, N., Novod, N., O'Neill, B., Osman, S., Markiewicz, E., Oyono, O.L., Patti, C.,  
755 Phunkhang, P., Pierre, F., Priest, M., Raghuraman, S., Rege, F., Reyes, R., Rise,  
756 C., Rogov, P., Ross, K., Ryan, E., Settipalli, S., Shea, T., Sherpa, N., Shi, L., Shih,  
757 D., Sparrow, T., Spaulding, J., Stalker, J., Stange-Thomann, N., Stavropoulos, S.,  
758 Stone, C., Strader, C., Tesfaye, S., Thomson, T., Thoulutsang, Y., Thoulutsang, D.,

Topham, K., Topping, I., Tsamla, T., Vassiliev, H., Vo, A., Wangchuk, T., Wangdi, T., Weiland, M., Wilkinson, J., Wilson, A., Yadav, S., Young, G., Yu, Q., Zembek, L., Zhong, D., Zimmer, A., Zwirko, Z., Alvarez, P., Brockman, W., Butler, J., Chin, C., Grabherr, M., Kleber, M., Mauceli, E., MacCallum, I., 2007. Evolution of genes and genomes on the Drosophila phylogeny. *Nature* 450, 203–18.  
doi:10.1038/nature06341

Couderc, J.-L., Godt, D., Zollman, S., Chen, J., Li, M., Tiong, S., Cramton, S.E., Sahut-Barnola, I., Laski, F.A., 2002. The bric à 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, 2419–33.

Davidson, E.H., 2006. *The Regulatory Genome: Gene Regulatory Networks In Development And Evolution*. Elsevier Inc., Burlington, MA.

Giraldo-Calderon, G.I., Emrich, S.J., MacCallum, R.M., Maslen, G., Emrich, S., Collins, F., Dialynas, E., Topalis, P., Ho, N., Gesing, S., Madey, G., Collins, F.H., Lawson, D., Kersey, P., Allen, J., Christensen, M., Hughes, D., Koscielny, G., Langridge, N., Gallego, E.L., Megy, K., Wilson, D., Gelbart, B., Emmert, D., Russo, S., Zhou, P., Christophides, G., Brockman, A., Kirmizoglou, I., MacCallum, B., Tiirikka, T., Louis, K., Dritsou, V., Mitra, E., Werner-Washburn, M., Baker, P., Platero, H., Aguilar, A., Bogol, S., Campbell, D., Carmichael, R., Cieslak, D., Davis, G., Konopinski, N., Nabrzyski, J., Reinking, C., Sheehan, A., Szakonyi, S., Wieck, R., 2015. VectorBase: An updated Bioinformatics Resource for invertebrate vectors and other organisms related with human diseases. *Nucleic Acids Res.* 43, D707–D713.

782       doi:10.1093/nar/gku1117

783   Gompel, N., Carroll, S.B., 2003. Genetic mechanisms and constraints governing the  
784       evolution of correlated traits in drosophilid flies. *Nature* 424, 931–935.  
785       doi:10.1038/nature01863.1.

786   Gompel, N., Prud'homme, B., Wittkopp, P.J., Kassner, V.A., Carroll, S.B., 2005. Chance  
787       caught on the wing: cis-regulatory evolution and the origin of pigment patterns in  
788       *Drosophila*. *Nature* 433, 481–7. doi:10.1038/nature03235

789   Groth, A.C., Fish, M., Nusse, R., Calos, M.P., 2004. Construction of Transgenic  
790       *Drosophila* by Using the Site-Specific Integrase From Phage  $\phi$ C31. *Genetics* 166,  
791       1775–1782.

792   Haley, B., Foys, B., Levine, M., 2010. Vectors and parameters that enhance the efficacy  
793       of RNAi-mediated gene disruption in transgenic *Drosophila*. *Proc. Natl. Acad. Sci.*  
794       U. S. A. 107, 11435–40. doi:10.1073/pnas.1006689107

795   Haley, B., Hendrix, D., Trang, V., Levine, M., 2008. A simplified miRNA-based gene  
796       silencing method for *Drosophila melanogaster*. *Dev. Biol.* 321, 482–90.  
797       doi:10.1016/j.ydbio.2008.06.015

798   Hare, E.E., Peterson, B.K., Iyer, V.N., Meier, R., Eisen, M.B., 2008. *Sepsid* even-  
799       skipped enhancers are functionally conserved in *Drosophila* despite lack of  
800       sequence conservation. *PLoS Genet.* 4, e1000106.  
801       doi:10.1371/journal.pgen.1000106

802   Hart, C.M., 2013. P{y-GAL4.G} insertions [WWW Document]. FlyBase. URL  
803       <http://flybase.org/reports/FBfr0221290.html>

804   Jeong, S., Rokas, A., Carroll, S.B., 2006. Regulation of body pigmentation by the

805 Abdominal-B Hox protein and its gain and loss in *Drosophila* evolution. *Cell* 125,  
806 1387–99. doi:10.1016/j.cell.2006.04.043

807 Kopp, A., Duncan, I., Carroll, S.B., 2000. Genetic control and evolution of sexually  
808 dimorphic characters in *Drosophila*. *Nature* 408, 553–9. doi:10.1038/35046017

809 Kriventseva, E. V., Tegenfeldt, F., Petty, T.J., Waterhouse, R.M., Simão, F.A.,  
810 Pozdnyakov, I.A., Ioannidis, P., Zdobnov, E.M., 2015. OrthoDB v8: Update of the  
811 hierarchical catalog of orthologs and the underlying free software. *Nucleic Acids*  
812 *Res.* 43, D250–D256. doi:10.1093/nar/gku1220

813 Levine, M., 2010. Transcriptional enhancers in animal development and evolution. *Curr.*  
814 *Biol.* 20, R754–63. doi:10.1016/j.cub.2010.06.070

815 Lours, C., Bardot, O., Godt, D., Laski, F.A., Couderc, J.-L., 2003. The *Drosophila*  
816 *melanogaster* BTB proteins bric a brac bind DNA through a composite DNA binding  
817 domain containing a pipsqueak and an AT-Hook motif. *Nucleic Acids Res.* 31,  
818 5389–5398. doi:10.1093/nar/gkg724

819 Ludwig, M.Z., Bergman, C., Patel, N.H., Kreitman, M., 2000. Evidence for stabilizing  
820 selection in a eukaryotic enhancer element. *Nature* 403, 564–7.  
821 doi:10.1038/35000615

822 Lynch, M., Conery, J.S., 2000. The evolutionary fate and consequences of duplicate  
823 genes. *Science* 290, 1151–5. doi:10.1126/science.290.5494.1151

824 Markstein, M., Pitsouli, C., Villalta, C., Celniker, S.E., Perrimon, N., 2008. Exploiting  
825 position effects and the gypsy retrovirus insulator to engineer precisely expressed  
826 transgenes. *Nat. Genet.* 40, 476–83. doi:10.1038/ng.101

827 Nitta, K.R., Jolma, A., Yin, Y., Morgunova, E., Kivioja, T., Akhtar, J., Hens, K., Toivonen,

J., Deplancke, B., Furlong, E.E.M., Taipale, J., 2015. Conservation of transcription factor binding specificities across 600 million years of bilateria evolution. *Elife* 4, 1–20. doi:10.7554/eLife.04837

Pfeiffer, B.D., Truman, J.W., Rubin, G.M., 2012. Using translational enhancers to increase transgene expression in *Drosophila*. *Proc. Natl. Acad. Sci. U. S. A.* 109, 6626–31. doi:10.1073/pnas.1204520109

Rebeiz, M., Williams, T.M., 2017. Using *Drosophila* pigmentation traits to study the mechanisms of cis-regulatory evolution. *Curr. Opin. Insect Sci.* 19, 1–7. doi:10.1016/j.cois.2016.10.002

Rebeiz, M., Williams, T.M., 2011. Experimental Approaches to Evaluate the Contributions of Candidate Cis- regulatory Mutations to Phenotypic Evolution., in: Orgogozo, V., Rockman, M. V. (Eds.), *Methods in Molecular Biology, Methods in Molecular Biology*. Humana Press, Totowa, NJ, pp. 351–375. doi:10.1007/978-1-61779-228-1

Richards, S., Liu, Y., Bettencourt, B.R., Hradecky, P., Letovsky, S., Nielsen, R., Thornton, K., Hubisz, M.J., Chen, R., Meisel, R.P., Couronne, O., Hua, S., Smith, M. a, Zhang, P., Liu, J., Bussemaker, H.J., van Batenburg, M.F., Howells, S.L., Scherer, S.E., Sodergren, E., Matthews, B.B., Crosby, M. a, Schroeder, A.J., Ortiz-Barrientos, D., Rives, C.M., Metzker, M.L., Muzny, D.M., Scott, G., Steffen, D., Wheeler, D. a, Worley, K.C., Havlak, P., Durbin, K.J., Egan, A., Gill, R., Hume, J., Morgan, M.B., Miner, G., Hamilton, C., Huang, Y., Waldron, L., Verduzco, D., Clerc-Blankenburg, K.P., Dubchak, I., Noor, M. a F., Anderson, W., White, K.P., Clark, A.G., Schaeffer, S.W., Gelbart, W., Weinstock, G.M., Gibbs, R. a, 2005.

Comparative genome sequencing of *Drosophila pseudoobscura*: chromosomal, gene, and cis-element evolution. *Genome Res.* 15, 1–18. doi:10.1101/gr.3059305

Rogers, W.A., Salomone, J.R., Tacy, D.J., Camino, E.M., Davis, K.A., Rebeiz, M., Williams, T.M., 2013. Recurrent Modification of a Conserved Cis-Regulatory Element Underlies Fruit Fly Pigmentation Diversity. *PLoS Genet.* 9, e1003740. doi:10.1371/journal.pgen.1003740

Rogers, W. a, Williams, T.M., 2011. Quantitative Comparison of *cis*-Regulatory Element (CRE) Activities in Transgenic *Drosophila melanogaster*. *J. Vis. Exp.* 2–7. doi:10.3791/3395

Salomone, J.R., Rogers, W.A., Rebeiz, M., Williams, T.M., 2013. The evolution of *Bab* paralog expression and abdominal pigmentation among *Sophophora* fruit fly species. *Evol. Dev.* 15, 442–57. doi:10.1111/ede.12053

Stern, D.L., Orgogozo, V., 2008. The loci of evolution: how predictable is genetic evolution? *Evolution* 62, 2155–77. doi:10.1111/j.1558-5646.2008.00450.x

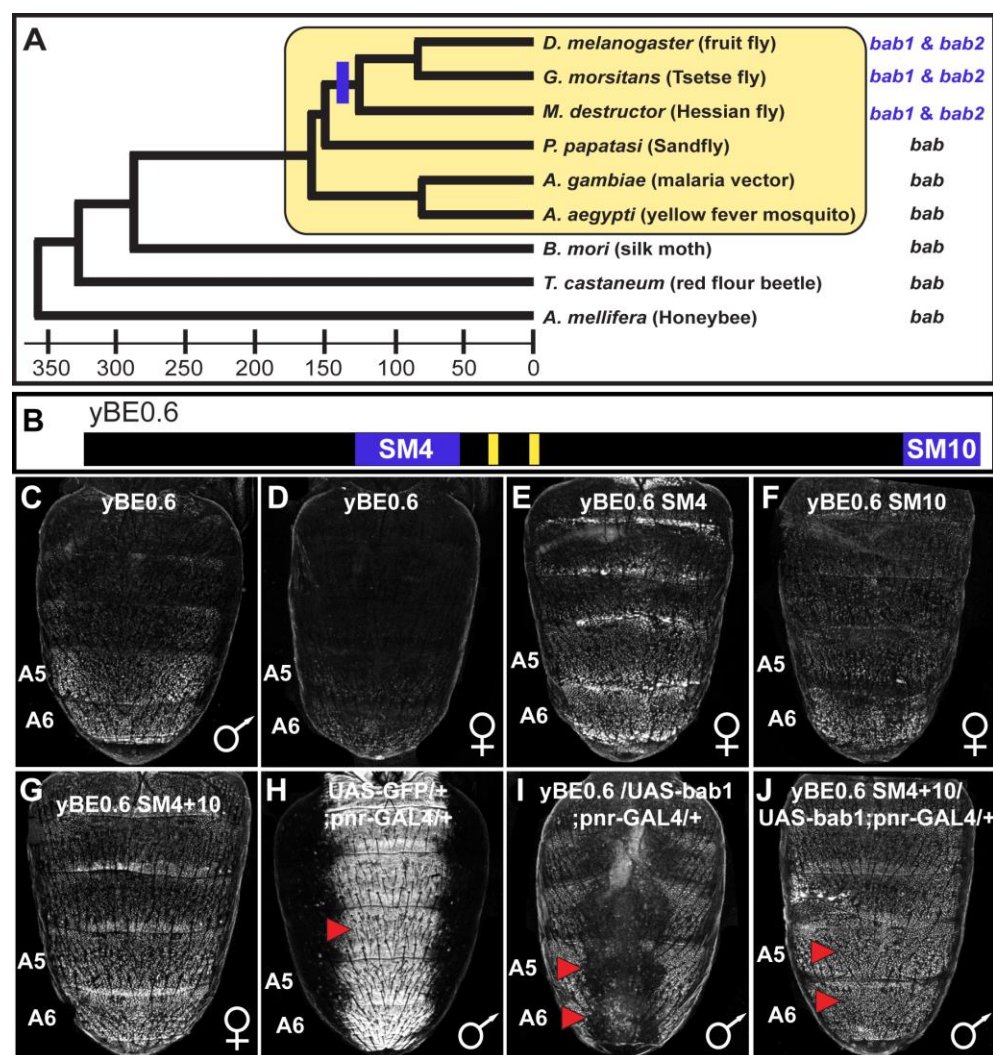
Swanson, C.I., Schwimmer, D.B., Barolo, S., 2011. Rapid Evolutionary Rewiring of a Structurally Constrained Eye Enhancer. *Curr. Biol.* 1–11. doi:10.1016/j.cub.2011.05.056

Vert, J., Foveau, N., Lajaunie, C., Vandenbrouck, Y., 2006. An accurate and interpretable model for siRNA efficacy prediction. *BMC Bioinformatics* 7, 1–17. doi:10.1186/1471-2105-7-520

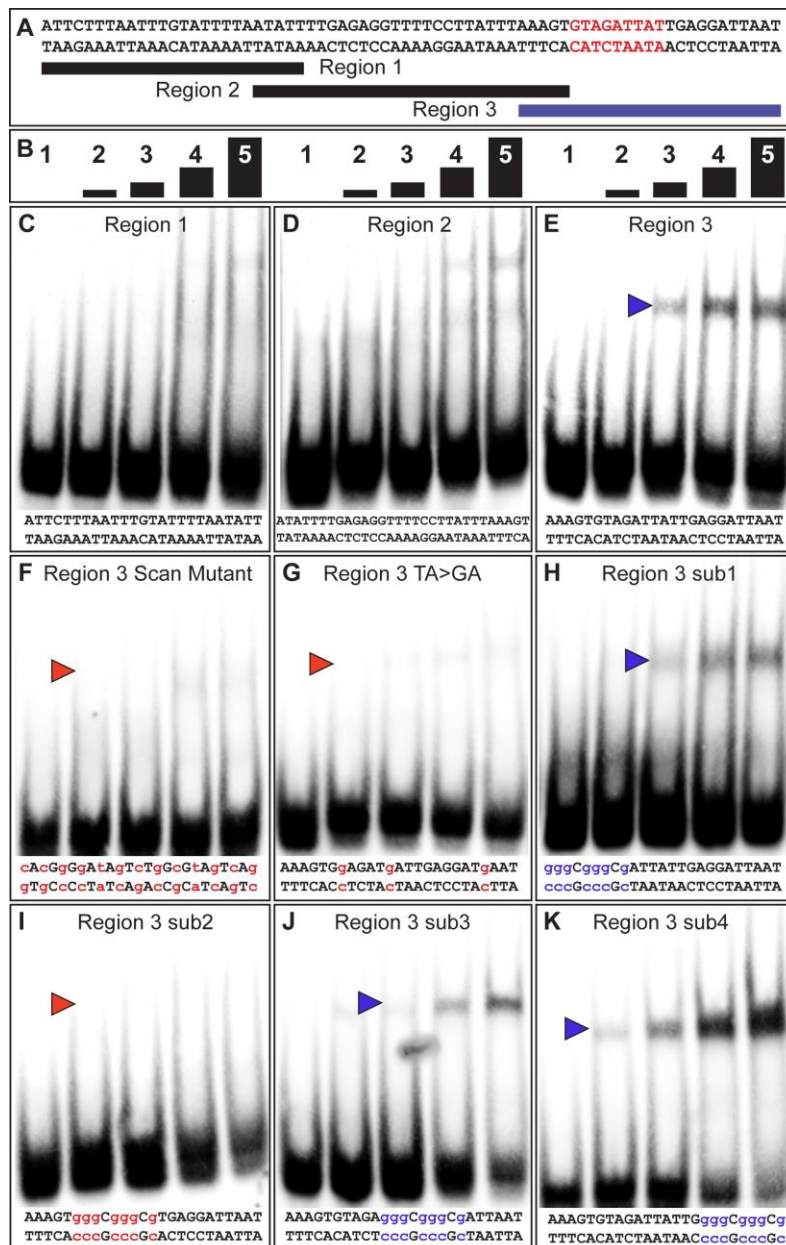
Wiegmann, B.M., Trautwein, M.D., Winkler, I.S., Barr, N.B., Kim, J., Blagoderov, V., Caravas, J., Narayanan, S., Schmidt-ott, U., Kampmeier, G.E., Meier, R., Yeates, D.K., 2011. Episodic radiations in the fly tree of life. *Proc Natl Acad Sci U S A* 108,

5690–5695. doi:10.1073/pnas.1012675108/-  
 /DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1012675108  
 Williams, J.A., Langeland, J.A., Thalley, B.S., Skeath, J.B., Carroll, S.B., 1995.  
 Expression of foreign proteins in E. coli using plasmid vectors and purification of  
 specific polyclonal antibodies, in: Glover, D.M., Hames, D.B. (Eds.), DNA Cloning 2  
 Expression Systems. Oxford University Press, Oxford, pp. 15–57.  
 Williams, T.M., Selegue, J.E., Werner, T., Gompel, N., Kopp, A., Carroll, S.B., 2008.  
 The regulation and evolution of a genetic switch controlling sexually dimorphic traits  
 in Drosophila. Cell 134, 610–23. doi:10.1016/j.cell.2008.06.052  
 Wittkopp, P.J., Vaccaro, K., Carroll, S.B., 2002. Evolution of yellow gene regulation and  
 pigmentation in Drosophila. Curr. Biol. 12, 1547–56.

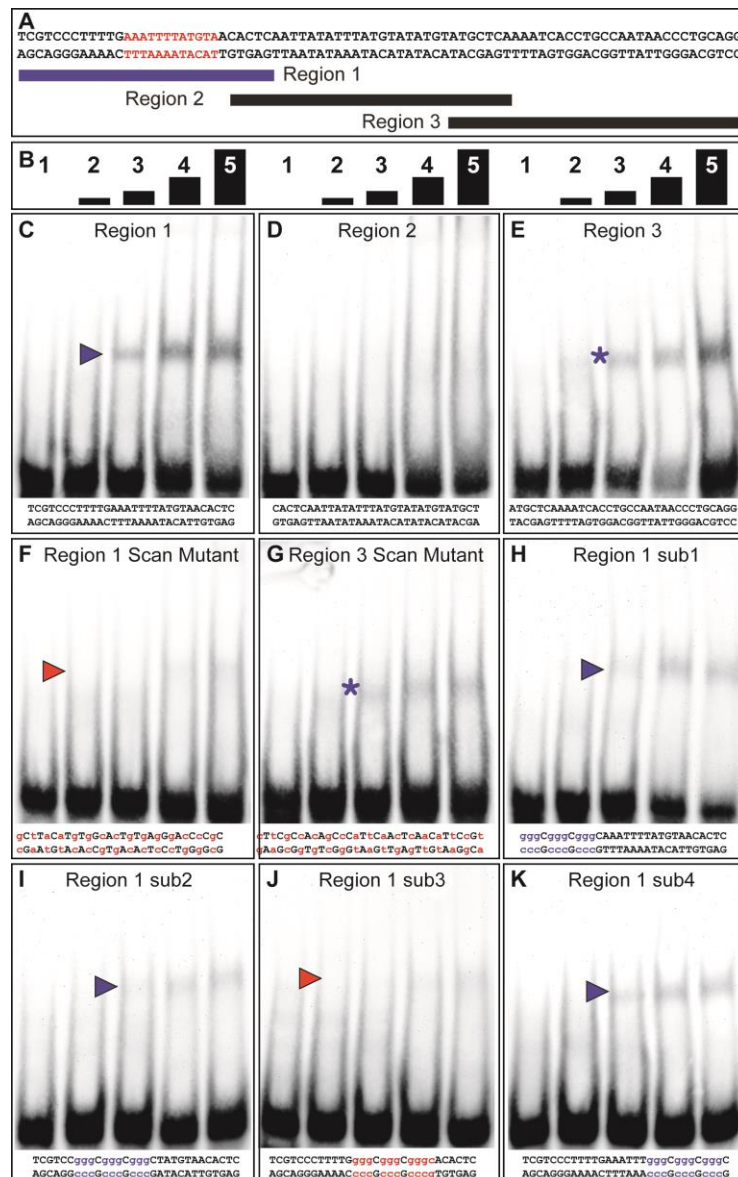
## Figures



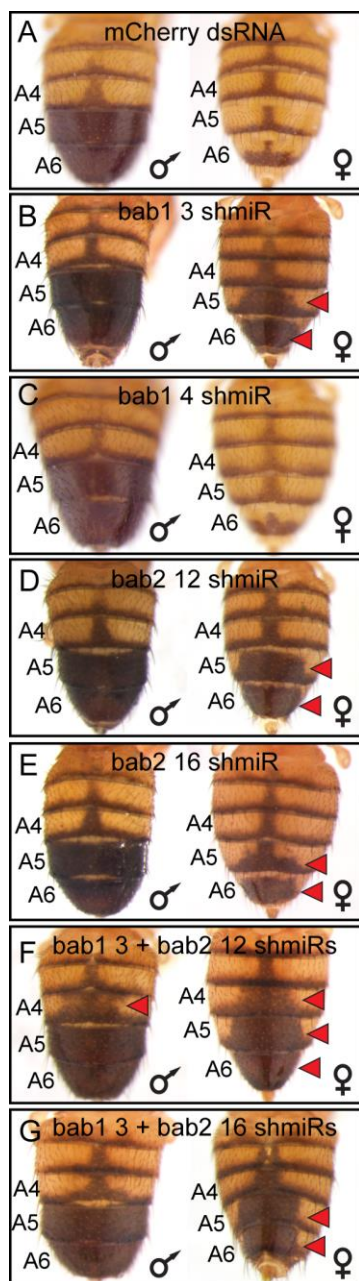
**Figure 1. The tandem duplicated *bab* genes perform a derived role in repressing a CRE controlling male-specific expression of the gene *yellow*.** (A) An ancestral *bab* gene was duplicated into the paralogous *bab1* and *bab2* genes in a Dipteran lineage that includes *Drosophila* fruit flies. The time scale indicates approximate divergence times in millions of years ago. (B) Male-specific expression of *yellow* in the abdominal epidermis is under the control of the yBE0.6 CRE that possesses two binding sites for Abd-B that are shown as yellow rectangles. Blue bars delimit the SM4 and SM10 regions required to suppress CRE activity in females. (C and D) The yBE0.6 EGFP reporter transgene is elevated in the male A5 and A6 abdomen segments (C) but is only barely detected females (D). (E-G) Ectopic reporter expression occurs in the female abdomen when either the SM4, SM10, or both regions are mutated. (H) The *pnr*-GAL4 driver activates dorsal midline expression of the UAS-EGFP gene, demarcating its domain of misexpression. (I) Dorsal midline expression of the yBE0.6 CRE is lost when *bab1* is ectopically expressed by *pnr*-GAL4. (J) When the SM4 and SM10 regions are mutated, the yBE0.6 CRE can activate reporter expression in midline regions in spite of ectopically expressed *bab1*.



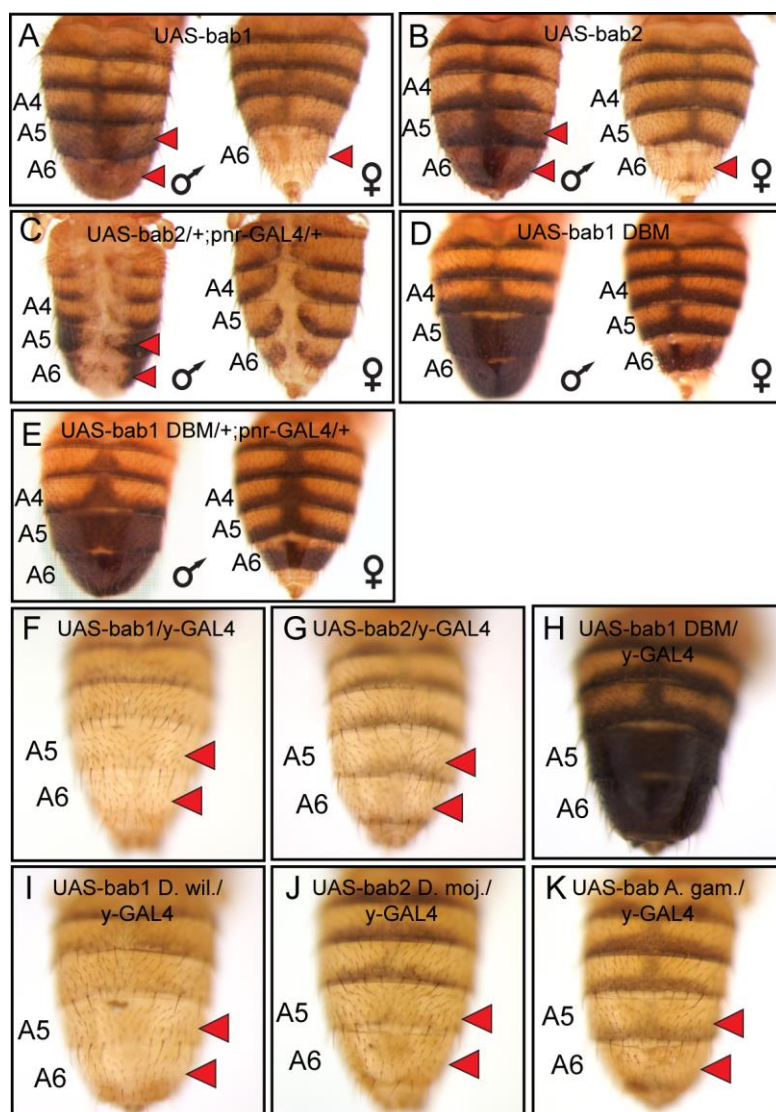
**Figure 2. The yBE0.6 possesses a binding site for Bab in the SM4 region.** (A) The wild type DNA sequence of the SM4 region is shown, which was subdivided into three smaller regions annotated below that were used as double stranded probes in gel shift assays with the GST-Bab1 DNA-binding Domain (Bab1-DBD). Red text delimits the inferred Bab-binding site. (B) Each probe was tested in gel shift assay reactions for binding with 5 different amounts of Bab1-DBD. These were from left to right: 0, 500, 1,000, 2,000, and 4,000 ng. (C-E) Gel shift assays using wild type probe sequences. (F-K) Gel shift assays using mutant probe sequences. Lower case blue letters indicate probe mutations that did not noticeably alter protein binding. Probe base pairs in lower case red letters are changes that altered protein binding. Blue and red arrowheads respectively indicate the location of shifted probe and where the quantity of shifted probe was noticeably reduced.



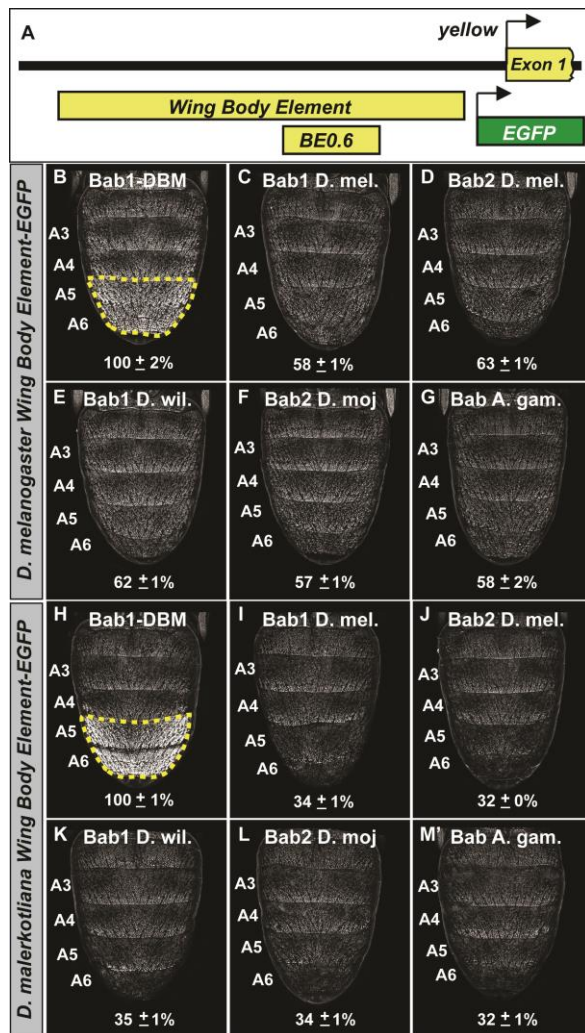
**Figure 3. The yBE0.6 possesses a binding site for Bab in the SM10 region.** (A) The wild type DNA sequence of the SM10 region is shown, which was subdivided into three smaller regions annotated below that were used as double stranded probes in gel shift assays with the GST-Bab1 DNA-binding Domain (Bab1-DBD). Red text delimits the inferred Bab-binding site. (B) Each probe was tested in gel shift assay reactions for binding with 5 different amounts of Bab1-DBD. These were from left to right: 0, 500, 1,000, 2,000, and 4,000 ng. (C-E). Gel shift assays using wild type probe sequences. (F-K) Gel shift assays using mutant probe sequences. Lower case purple letters indicate probe mutations that did not noticeably alter protein binding. Probe base pairs in lower case red letters are changes that altered protein binding. Purple and red arrowheads respectively indicate the location of shifted probe and where the quantity of shifted probe was noticeably reduced. Asterisks indicate a situation where binding was non-specific as both the wild type and mutant probes were bound by the Bab1-DBD.



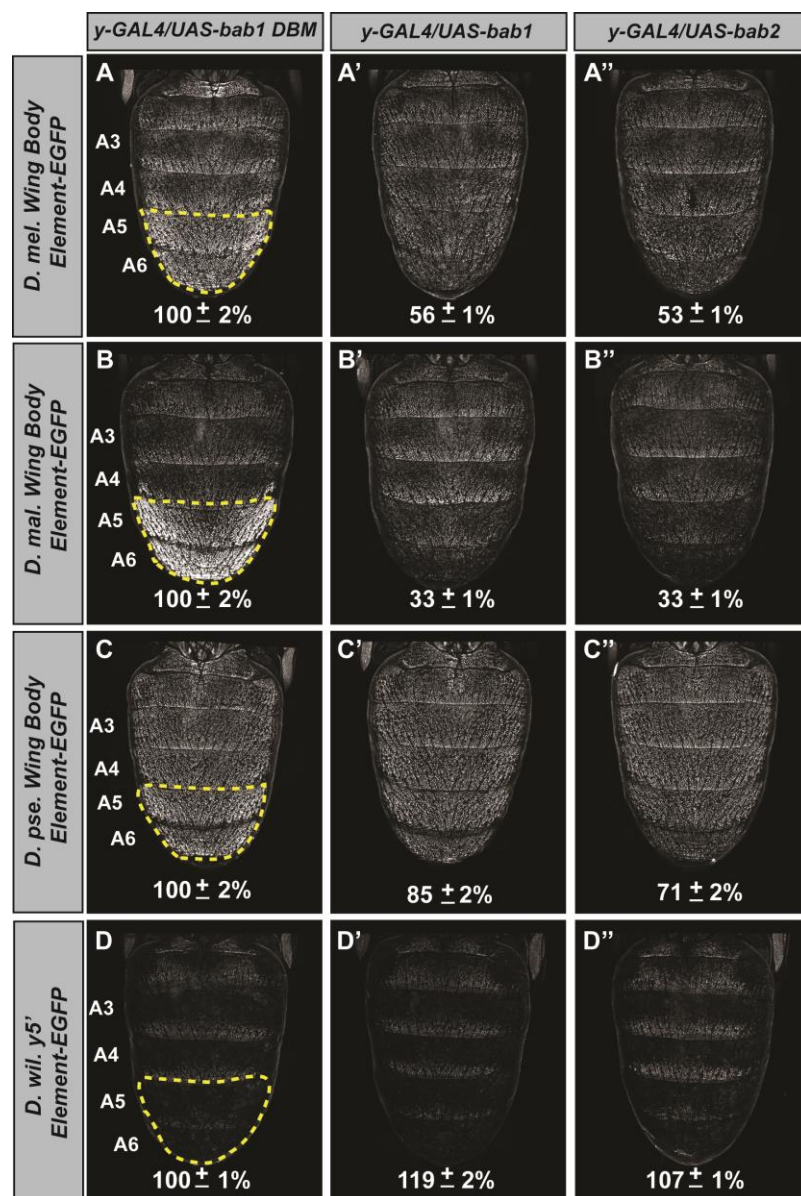
**Figure 4. RNA-interference reveals a necessity for both *bab1* and *bab2* in suppressing female tergite pigmentation.** (A-G) Double stranded (ds) RNA transgenes with UAS binding sites were expressed in the dorsal midline abdomen region driven by GAL4 that was expressed in the midline pattern of the *pnr* gene. (A) Expression of a negative control dsRNA that targets a gene (*mCherry*) that does not naturally exist in the *D. melanogaster* genome resulted in no apparent pigmentation phenotype from RNA-interference (RNA-i). (B and C) Two different dsRNAs specific to *bab1* and to (F and G) *bab2* were tested for pigmentation phenotypes from RNA-i. (F and G) Simultaneous RNA-i for *bab1* and *bab2* was accomplished by expressing “chained” transgenes. Red arrowheads indicate tergite regions where RNA-i caused the development of ectopic pigmentation.



**Figure 5. Bab1 and Bab2 are sufficient to suppress tergite pigmentation as DNA-binding transcription factors.** (A-K) Ectopic expression assays for the protein coding sequence of *D. melanogaster* *bab1*, *bab2*, and a DNA-binding compromised version of *bab1* (*bab1*-DBM). (A, B, and D) Leaky expression of transgenes from the attP40 transgene insertion site. (C and E) Ectopic expression of protein coding sequences in the dorsal midline of male and female abdomens driven by *pnr*-GAL4. (F-K) Ectopic expression of protein coding sequences in the male abdomen under the control of the *y*-GAL4 transgene. Red arrowheads indicate tergite regions with conspicuously reduced tergite pigmentation.

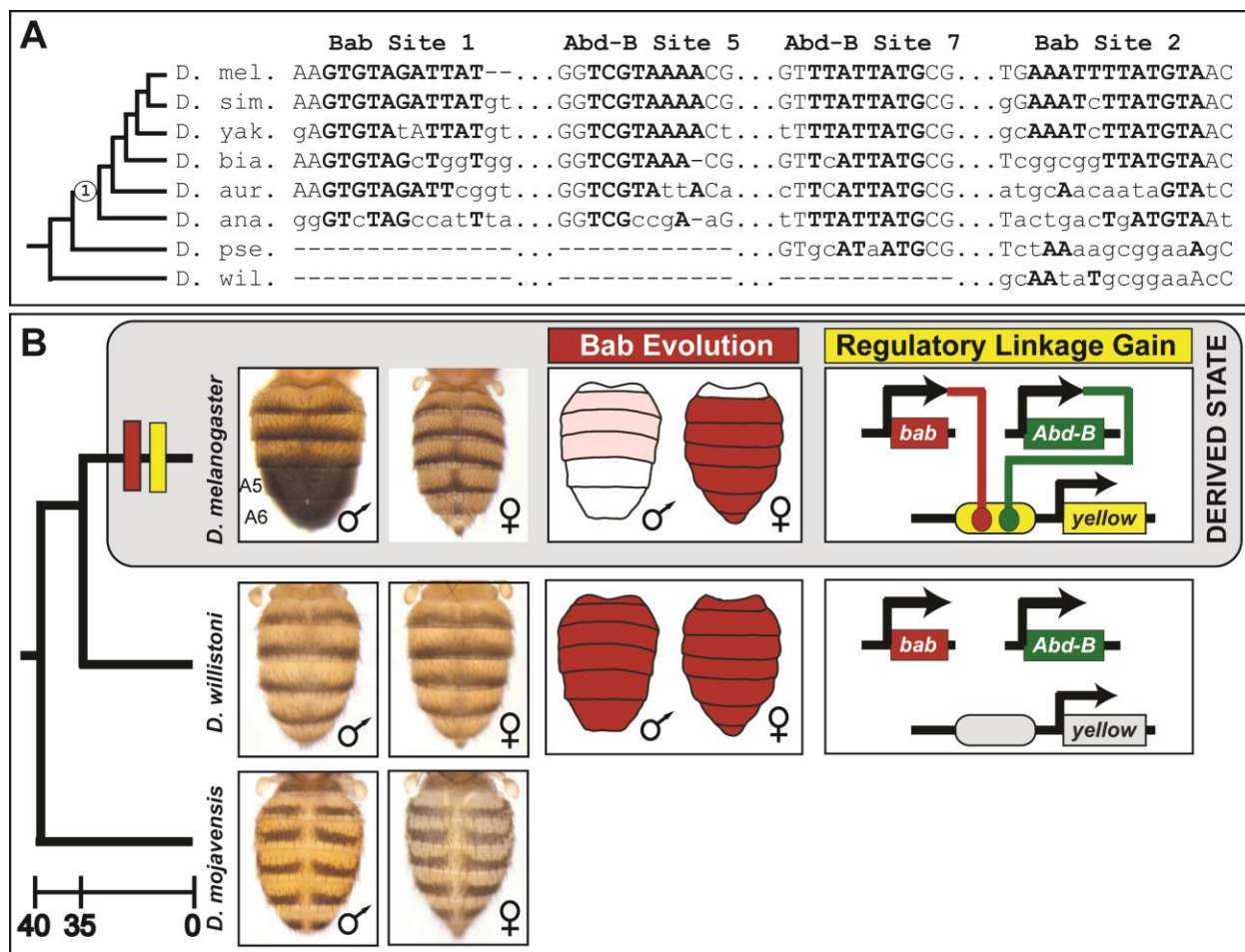


**Figure 6. The Bab paralogs can suppress the male-specific activity of the regulatory region containing the wing element and body element CREs.** (A) 5' of yellow exon 1 resides the wing element and body element CREs, and the position of the *D. melanogaster* yBE0.6 is shown below the to-scale representation of the partial locus. (B-G) Comparison of the levels of EGFP-reporter expression in the male A5 and A6 segments driven by the Wing Body Element of *D. melanogaster*. (H-M) Comparison of the levels of EGFP-reporter expression in the male A5 and A6 segments driven by the Wing Body Element of *D. malerkotliana*. The levels of EGFP expression are represented as the % of the mean ± SEM for samples in which the Bab1-DBM was expressed. (B and H) Robust EGFP reporter expression in samples ectopically expressing the Bab1-DBM protein in the *y*-GAL4 pattern. (C and I) Ectopic expression of Bab1 in the *y*-GAL4 pattern reduced A5 and A6 expression compared to the control. (D and J) Ectopic expression of Bab2 in the *y*-GAL4 pattern reduced A5 and A6 expression compared to the control. (E and K) Ectopic expression of *D. willistoni* Bab1 in the *y*-GAL4 pattern reduced A5 and A6 expression compared to the control. (F and L) Ectopic expression of *D. mojavensis* Bab2 in the *y*-GAL4 pattern reduced A5 and A6 expression compared to the control. (G and M) Ectopic expression of *A. gambiae* Bab in the *y*-GAL4 pattern reduced A5 and A6 expression compared to the control.



**Figure 7. The evolved repression by Bab for *cis*-regulatory regions 5' of *yellow*.**

(A-A'') Comparison of the levels of EGFP-reporter expression in the male A5 and A6 segments driven by the Wing Body Element of *D. melanogaster*. (B-B'') Comparison of the levels of EGFP-reporter expression in the male A5 and A6 segments driven by the Wing Body Element of *D. malerkotliana*. (C-C'') Comparison of the levels of EGFP-reporter expression in the male A5 and A6 segments driven by the Wing Body Element of *D. pseudoobscura*. (D-D'') Comparison of the levels of EGFP-reporter expression in the male A5 and A6 segments driven by the 5' non-coding region of *D. willistoni yellow*. For each comparison, the level of EGFP expression are expressed as the percentage of the mean  $\pm$  SEM for samples in which the Bab1-DBM was expressed. (A-D) Ectopic expression of the Bab1-DBM in the y-GAL4 pattern. (A'-D') Ectopic expression of Bab1 in the y-GAL4 pattern. (A''-D'') Ectopic expression of Bab2 in the y-GAL4 pattern.



**Figure 8. The evolution of male-specific pigmentation required the gain of a regulatory linkage between Bab and the newly evolved body element CRE controlling *yellow* expression.** (A) An alignment of the Bab-bound sequences in the SM4 (site 1) and SM10 (site 2) regions and for the two previously identified binding sites for Abd-B in the yBE0.6 CRE (Jeong et al. 2006). “Node 1” on the phylogeny indicates the most recent common ancestor suspected to have possessed the derived male-specific pattern of pigmentation. Time scale shown is in millions of years ago. Bold capital letters indicate the bases bound by the transcription factor in the *D. melanogaster* CRE, and those which are conserved in the orthologous regions for related species. (B) Model for the derivation of a dimorphic pigmentation trait where dimorphic pigmentation required the evolution of a dimorphic Bab expression and the gain of a regulatory linkage between Bab and *yellow* through gains of binding sites in the body element CRE.

## 1022 Supplementary Figures

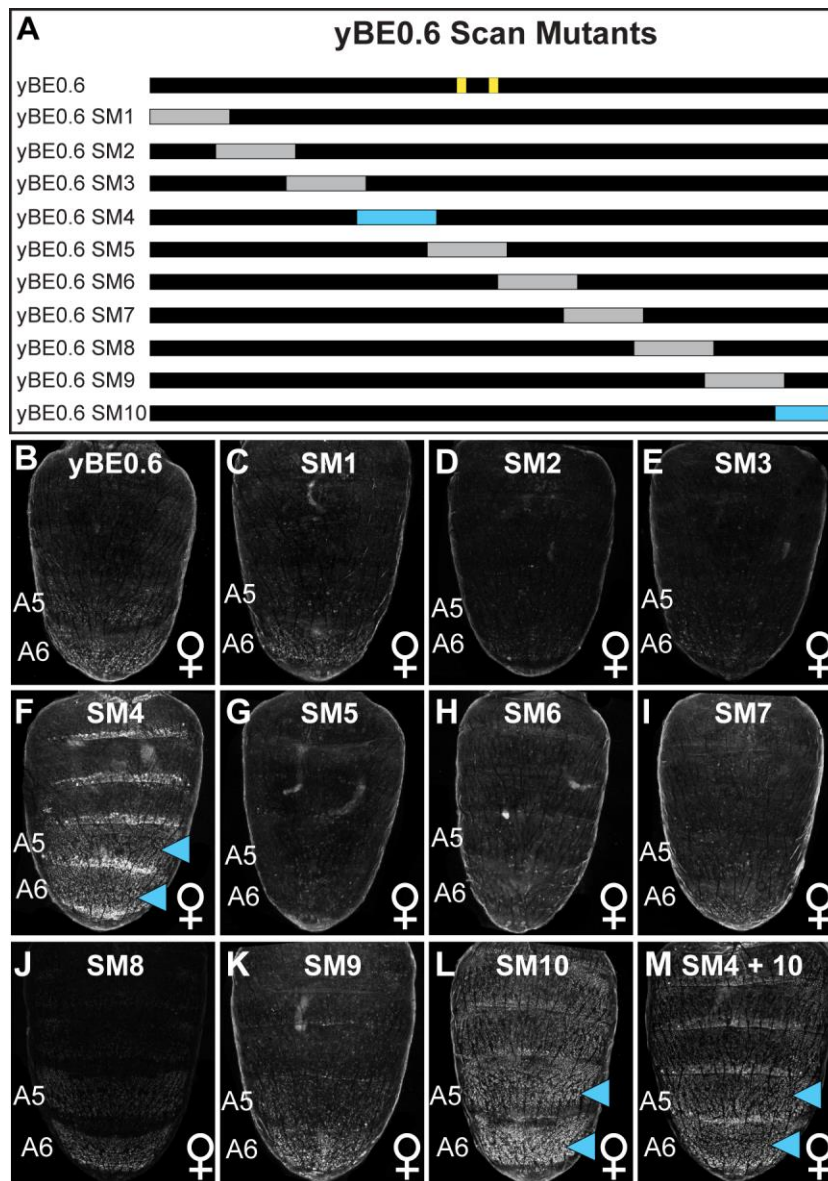
1023			<b>BE2.5 Fwd</b>				
1024	yBE0.6	1	GGCGCGCCCT	GTGGGTGCAA	TGATTTAGAA	TGCGGGCAAG	GGATCAAGTT
1025	SM1	1	GGCGCGCCaT	tTtGtTtCcA	gGcTgTcGcA	gGaGtGaAcG	tGcTaAcGgT
1026	SM2	1	GGCGCGCCCT	GTGGGTGCAA	TGATTTAGAA	TGCGGGCAAG	GGATCAAGTT
1027	SM3	1	GGCGCGCCCT	GTGGGTGCAA	TGATTTAGAA	TGCGGGCAAG	GGATCAAGTT
1028	SM4	1	GGCGCGCCCT	GTGGGTGCAA	TGATTTAGAA	TGCGGGCAAG	GGATCAAGTT
1029	SM5	1	GGCGCGCCCT	GTGGGTGCAA	TGATTTAGAA	TGCGGGCAAG	GGATCAAGTT
1030	SM6	1	GGCGCGCCCT	GTGGGTGCAA	TGATTTAGAA	TGCGGGCAAG	GGATCAAGTT
1031	SM7	1	GGCGCGCCCT	GTGGGTGCAA	TGATTTAGAA	TGCGGGCAAG	GGATCAAGTT
1032	SM8	1	GGCGCGCCCT	GTGGGTGCAA	TGATTTAGAA	TGCGGGCAAG	GGATCAAGTT
1033	SM9	1	GGCGCGCCCT	GTGGGTGCAA	TGATTTAGAA	TGCGGGCAAG	GGATCAAGTT
1034	SM10	1	GGCGCGCCCT	GTGGGTGCAA	TGATTTAGAA	TGCGGGCAAG	GGATCAAGTT
1035							
1036							
1037	yBE0.6	51	GAACCACTTC	TAAGAAAAAA	TAGCATTGCA	TAAATGATAT	AGAGTCCAAA
1038	SM1	51	tAcCaAaTgC	gAcGcAcAcA	gAtCcTgGCA	TAAATGATAT	AGAGTCCAAA
1039	SM2	51	GAACCACTTC	TAAGAAAAcA	gAtCcTgGaA	gAcAgGcTcT	cGcGgCaAcA
1040	SM3	51	GAACCACTTC	TAAGAAAAAA	TAGCATTGCA	TAAATGATAT	AGAGTCCAAA
1041	SM4	51	GAACCACTTC	TAAGAAAAAA	TAGCATTGCA	TAAATGATAT	AGAGTCCAAA
1042	SM5	51	GAACCACTTC	TAAGAAAAAA	TAGCATTGCA	TAAATGATAT	AGAGTCCAAA
1043	SM6	51	GAACCACTTC	TAAGAAAAAA	TAGCATTGCA	TAAATGATAT	AGAGTCCAAA
1044	SM7	51	GAACCACTTC	TAAGAAAAAA	TAGCATTGCA	TAAATGATAT	AGAGTCCAAA
1045	SM8	51	GAACCACTTC	TAAGAAAAAA	TAGCATTGCA	TAAATGATAT	AGAGTCCAAA
1046	SM9	51	GAACCACTTC	TAAGAAAAAA	TAGCATTGCA	TAAATGATAT	AGAGTCCAAA
1047	SM10	51	GAACCACTTC	TAAGAAAAAA	TAGCATTGCA	TAAATGATAT	AGAGTCCAAA
1048							
1049							
1050	yBE0.6	101	AACTACACAA	ATTCAATAGC	AGTAATGGTT	ACATTAGCTT	TGAAATTGTT
1051	SM1	101	AACTACACAA	ATTCAATAGC	AGTAATGGTT	ACATTAGCTT	TGAAATTGTT
1052	SM2	101	cAaTcCcCcA	cTgCcAgAtC	cGgAcTtGgT	cCcTgAtCTT	TGAAATTGTT
1053	SM3	101	AACTACACAA	ATTCAATAGC	AGTAATGGgT	cCcTgAtCgT	gGcAcTgGgT
1054	SM4	101	AACTACACAA	ATTCAATAGC	AGTAATGGTT	ACATTAGCTT	TGAAATTGTT
1055	SM5	101	AACTACACAA	ATTCAATAGC	AGTAATGGTT	ACATTAGCTT	TGAAATTGTT
1056	SM6	101	AACTACACAA	ATTCAATAGC	AGTAATGGTT	ACATTAGCTT	TGAAATTGTT
1057	SM7	101	AACTACACAA	ATTCAATAGC	AGTAATGGTT	ACATTAGCTT	TGAAATTGTT
1058	SM8	101	AACTACACAA	ATTCAATAGC	AGTAATGGTT	ACATTAGCTT	TGAAATTGTT
1059	SM9	101	AACTACACAA	ATTCAATAGC	AGTAATGGTT	ACATTAGCTT	TGAAATTGTT
1060	SM10	101	AACTACACAA	ATTCAATAGC	AGTAATGGTT	ACATTAGCTT	TGAAATTGTT
1061							
1062							
1063	yBE0.6	151	TTTAGACATC	CGAAGAAATA	AGATTAAATT	TAAACGGCAT	TCTTTAATTT
1064	SM1	151	TTTAGACATC	CGAAGAAATA	AGATTAAATT	TAAACGGCAT	TCTTTAATTT
1065	SM2	151	TTTAGACATC	CGAAGAAATA	AGATTAAATT	TAAACGGCAT	TCTTTAATTT
1066	SM3	151	gTgAtAaAgC	aGcAtAcAgA	cGcTgAcAgT	gAcAaGtCcT	gCgTgAcTTT
1067	SM4	151	TTTAGACATC	CGAAGAAATA	AGATTAAATT	TAAACGGCcT	gCgTgAcTgT
1068	SM5	151	TTTAGACATC	CGAAGAAATA	AGATTAAATT	TAAACGGCAT	TCTTTAATTT
1069	SM6	151	TTTAGACATC	CGAAGAAATA	AGATTAAATT	TAAACGGCAT	TCTTTAATTT
1070	SM7	151	TTTAGACATC	CGAAGAAATA	AGATTAAATT	TAAACGGCAT	TCTTTAATTT
1071	SM8	151	TTTAGACATC	CGAAGAAATA	AGATTAAATT	TAAACGGCAT	TCTTTAATTT
1072	SM9	151	TTTAGACATC	CGAAGAAATA	AGATTAAATT	TAAACGGCAT	TCTTTAATTT
1073	SM10	151	TTTAGACATC	CGAAGAAATA	AGATTAAATT	TAAACGGCAT	TCTTTAATTT
1074							
1075						<b>Bab1 bound region</b>	
1076	yBE0.6	201	GTATTTTAAAT	ATTTTGAAGAG	GTTTTTCCTTA	TTTAAAGTGT	AGATTATTGA
1077	SM1	201	GTATTTTAAAT	ATTTTGAAGAG	GTTTTTCCTTA	TTTAAAGTGT	AGATTATTGA

1078	SM2	201	GTATTTTAAT	ATTTTGAGAG	GTTTTCCTTA	TTTAAAGTGT	AGATTATTGA
1079	SM3	201	GTATTTTAAT	ATTTTGAGAG	GTTTTCCTTA	TTTAAAGTGT	AGATTATTGA
1080	SM4	201	tTcTgTgAcT	cTgTgGcGcG	tTgTgCaTgA	gTgAcAtTtT	cGcTgAgTtA
1081	SM5	201	GTATTTTAAT	ATTTTGAGAG	GTTTTCCTTA	TTTAAAGTGT	AGATTATTtA
1082	SM6	201	GTATTTTAAT	ATTTTGAGAG	GTTTTCCTTA	TTTAAAGTGT	AGATTATTGA
1083	SM7	201	GTATTTTAAT	ATTTTGAGAG	GTTTTCCTTA	TTTAAAGTGT	AGATTATTGA
1084	SM8	201	GTATTTTAAT	ATTTTGAGAG	GTTTTCCTTA	TTTAAAGTGT	AGATTATTGA
1085	SM9	201	GTATTTTAAT	ATTTTGAGAG	GTTTTCCTTA	TTTAAAGTGT	AGATTATTGA
1086	SM10	201	GTATTTTAAT	ATTTTGAGAG	GTTTTCCTTA	TTTAAAGTGT	AGATTATTGA
1087							
1088							
1089	yBE0.6	251	GGATTAATGC	AAACCACCTTT	ATCTGCGGAG	GTCGTAAAAC	GTATTTTTTAC
1090	SM1	251	GGATTAATGC	AAACCACCTTT	ATCTGCGGAG	GTCGTAAAAC	GTATTTTTTAC
1091	SM2	251	GGATTAATGC	AAACCACCTTT	ATCTGCGGAG	GTCGTAAAAC	GTATTTTTTAC
1092	SM3	251	GGATTAATGC	AAACCACCTTT	ATCTGCGGAG	GTCGTAAAAC	GTATTTTTTAC
1093	SM4	251	tGcTgAcTGC	AAACCACCTTT	ATCTGCGGAG	GTCGTAAAAC	GTATTTTTTAC
1094	SM5	251	tGcTgAcTtC	cAcCaAaTgT	cTaTtCtGcG	tTCGTAAAAa	GgAgTgTgAa
1095	SM6	251	GGATTAATGC	AAACCACCTTT	ATCTGCGGAG	GTCGTAAAAC	GTATTTTTTAC
1096	SM7	251	GGATTAATGC	AAACCACCTTT	ATCTGCGGAG	GTCGTAAAAC	GTATTTTTTAC
1097	SM8	251	GGATTAATGC	AAACCACCTTT	ATCTGCGGAG	GTCGTAAAAC	GTATTTTTTAC
1098	SM9	251	GGATTAATGC	AAACCACCTTT	ATCTGCGGAG	GTCGTAAAAC	GTATTTTTTAC
1099	SM10	251	GGATTAATGC	AAACCACCTTT	ATCTGCGGAG	GTCGTAAAAC	GTATTTTTTAC
1100							
1101							
1102	yBE0.6	301	CCATTTGCAT	GTTTATTATG	CGTGTGGCTG	GTTGTATTAC	TTTACTTAAG
1103	SM1	301	CCATTTGCAT	GTTTATTATG	CGTGTGGCTG	GTTGTATTAC	TTTACTTAAG
1104	SM2	301	CCATTTGCAT	GTTTATTATG	CGTGTGGCTG	GTTGTATTAC	TTTACTTAAG
1105	SM3	301	CCATTTGCAT	GTTTATTATG	CGTGTGGCTG	GTTGTATTAC	TTTACTTAAG
1106	SM4	301	CCATTTGCAT	GTTTATTATG	CGTGTGGCTG	GTTGTATTAC	TTTACTTAAG
1107	SM5	301	CaAgTgGaAg	GgTgAgTcTG	CGTGTGGCTG	GTTGTATTAC	TTTACTTAAG
1108	SM6	301	CCATTTGCcT	gTtTATTATG	aGgGgGtCgG	tTgGgAgTcC	gTgAaTgAcG
1109	SM7	301	CCATTTGCAT	GTTTATTATG	CGTGTGGCTG	GTTGTATTAC	TTTACTTAAG
1110	SM8	301	CCATTTGCAT	GTTTATTATG	CGTGTGGCTG	GTTGTATTAC	TTTACTTAAG
1111	SM9	301	CCATTTGCAT	GTTTATTATG	CGTGTGGCTG	GTTGTATTAC	TTTACTTAAG
1112	SM10	301	CCATTTGCAT	GTTTATTATG	CGTGTGGCTG	GTTGTATTAC	TTTACTTAAG
1113							
1114	yBE0.6	351	TTTTGCAATT	TTTTCTTTAG	CAAGCAGGTG	CATTTGGGCC	AAGAGATATA
1115	SM1	351	TTTTGCAATT	TTTTCTTTAG	CAAGCAGGTG	CATTTGGGCC	AAGAGATATA
1116	SM2	351	TTTTGCAATT	TTTTCTTTAG	CAAGCAGGTG	CATTTGGGCC	AAGAGATATA
1117	SM3	351	TTTTGCAATT	TTTTCTTTAG	CAAGCAGGTG	CATTTGGGCC	AAGAGATATA
1118	SM4	351	TTTTGCAATT	TTTTCTTTAG	CAAGCAGGTG	CATTTGGGCC	AAGAGATATA
1119	SM5	351	TTTTGCAATT	TTTTCTTTAG	CAAGCAGGTG	CATTTGGGCC	AAGAGATATA
1120	SM6	351	gTgTtCcAgT	gTgTaTgTcG	aAcGaAtGTG	CATTTGGGCC	AAGAGATATA
1121	SM7	351	TTTTGCAATT	TTTTCTTTaG	aAcGaAtGgG	aAgTgGtGaC	cAtAtAgAgA
1122	SM8	351	TTTTGCAATT	TTTTCTTTAG	CAAGCAGGTG	CATTTGGGCC	AAGAGATATA
1123	SM9	351	TTTTGCAATT	TTTTCTTTAG	CAAGCAGGTG	CATTTGGGCC	AAGAGATATA
1124	SM10	351	TTTTGCAATT	TTTTCTTTAG	CAAGCAGGTG	CATTTGGGCC	AAGAGATATA
1125							
1126							
1127	yBE0.6	401	TGCGATCGCT	TTCGGTTTCGA	ATTTTTTAACA	TTTACTTGCG	GCGATGGTCA
1128	SM1	401	TGCGATCGCT	TTCGGTTTCGA	ATTTTTTAACA	TTTACTTGCG	GCGATGGTCA
1129	SM2	401	TGCGATCGCT	TTCGGTTTCGA	ATTTTTTAACA	TTTACTTGCG	GCGATGGTCA
1130	SM3	401	TGCGATCGCT	TTCGGTTTCGA	ATTTTTTAACA	TTTACTTGCG	GCGATGGTCA
1131	SM4	401	TGCGATCGCT	TTCGGTTTCGA	ATTTTTTAACA	TTTACTTGCG	GCGATGGTCA
1132	SM5	401	TGCGATCGCT	TTCGGTTTCGA	ATTTTTTAACA	TTTACTTGCG	GCGATGGTCA
1133	SM6	401	TGCGATCGCT	TTCGGTTTCGA	ATTTTTTAACA	TTTACTTGCG	GCGATGGTCA
1134	SM7	401	gGaGcTaGaT	gTaGtTgCtA	cTgTgtcAaA	gTgAaTgGCG	GCGATGGTCA

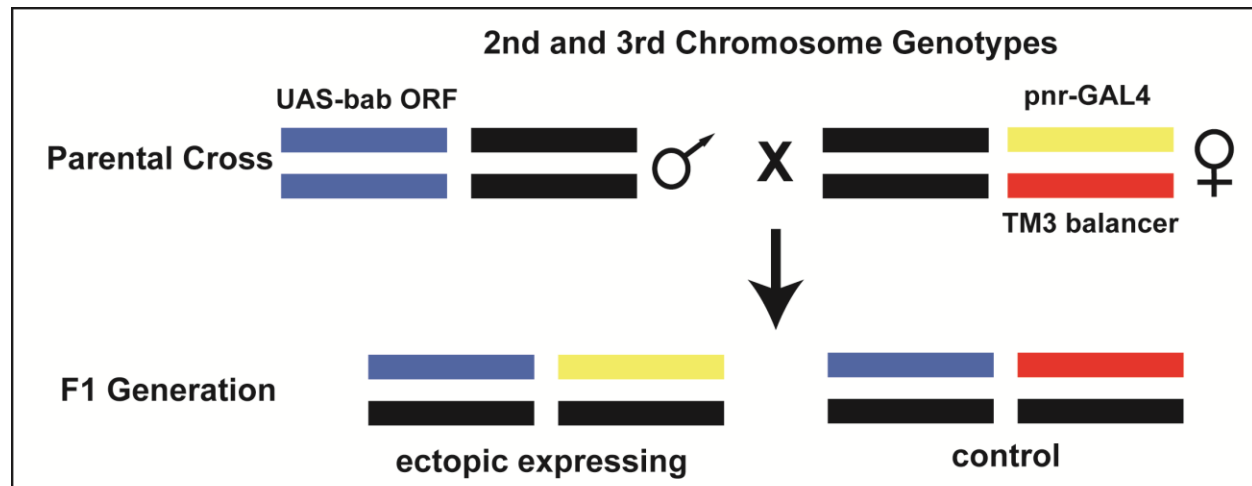
1135	SM8	401	TGCGATCGCT	TTCGGTTCGA	ATTTTTAAaA	gTgAaTgGaG	tCtAgGtTaA
1136	SM9	401	TGCGATCGCT	TTCGGTTCGA	ATTTTTAACA	TTTACTTGCG	GCGATGGTCA
1137	SM10	401	TGCGATCGCT	TTCGGTTCGA	ATTTTTAACA	TTTACTTGCG	GCGATGGTCA
1138							
1139							
1140	yBE0.6	451	TTAGAGCATT	ACCCACTTAG	GGCACCCCCA	ACATCCAGTT	GATTTTCAGG
1141	SM1	451	TTAGAGCATT	ACCCACTTAG	GGCACCCCCA	ACATCCAGTT	GATTTTCAGG
1142	SM2	451	TTAGAGCATT	ACCCACTTAG	GGCACCCCCA	ACATCCAGTT	GATTTTCAGG
1143	SM3	451	TTAGAGCATT	ACCCACTTAG	GGCACCCCCA	ACATCCAGTT	GATTTTCAGG
1144	SM4	451	TTAGAGCATT	ACCCACTTAG	GGCACCCCCA	ACATCCAGTT	GATTTTCAGG
1145	SM5	451	TTAGAGCATT	ACCCACTTAG	GGCACCCCCA	ACATCCAGTT	GATTTTCAGG
1146	SM6	451	TTAGAGCATT	ACCCACTTAG	GGCACCCCCA	ACATCCAGTT	GATTTTCAGG
1147	SM7	451	TTAGAGCATT	ACCCACTTAG	GGCACCCCCA	ACATCCAGTT	GATTTTCAGG
1148	SM8	451	gTcGcGaAgT	cCaCcGtGc	tGaAaCaCaA	cCcTaCcGgT	tAgTgTaAGG
1149	SM9	451	TTAGAGCATT	ACCCACTTAG	GGCACCCCCA	ACATCCAGgT	tAgTgTaAtG
1150	SM10	451	TTAGAGCATT	ACCCACTTAG	GGCACCCCCA	ACATCCAGTT	GATTTTCAGG
1151							
1152							
1153	yBE0.6	501	GACCACAATA	TTTTAAATAA	CAGCTAGTGG	AATTACCTAA	AAGCGCTTTC
1154	SM1	501	GACCACAATA	TTTTAAATAA	CAGCTAGTGG	AATTACCTAA	AAGCGCTTTC
1155	SM2	501	GACCACAATA	TTTTAAATAA	CAGCTAGTGG	AATTACCTAA	AAGCGCTTTC
1156	SM3	501	GACCACAATA	TTTTAAATAA	CAGCTAGTGG	AATTACCTAA	AAGCGCTTTC
1157	SM4	501	GACCACAATA	TTTTAAATAA	CAGCTAGTGG	AATTACCTAA	AAGCGCTTTC
1158	SM5	501	GACCACAATA	TTTTAAATAA	CAGCTAGTGG	AATTACCTAA	AAGCGCTTTC
1159	SM6	501	GACCACAATA	TTTTAAATAA	CAGCTAGTGG	AATTACCTAA	AAGCGCTTTC
1160	SM7	501	GACCACAATA	TTTTAAATAA	CAGCTAGTGG	AATTACCTAA	AAGCGCTTTC
1161	SM8	501	GACCACAATA	TTTTAAATAA	CAGCTAGTGG	AATTACCTAA	AAGCGCTTTC
1162	SM9	501	tAaCcCcAgA	gTgTcAcTcA	aAtCgAtTtG	cAgTcCaTcA	cAtCtCgTgC
1163	SM10	501	GACCACAATA	TTTTAAATAA	CAGCTAGTGG	AATTACCTAA	AAGCGCTTgC
1164							
1165			<b>Bab1 bound region</b>				
1166	yBE0.6	551	GTCCCTTTTG	AAATTTTATG	TAACACTCAA	TTATATTTAT	GTATATGTAT
1167	SM1	551	GTCCCTTTTG	AAATTTTATG	TAACACTCAA	TTATATTTAT	GTATATGTAT
1168	SM2	551	GTCCCTTTTG	AAATTTTATG	TAACACTCAA	TTATATTTAT	GTATATGTAT
1169	SM3	551	GTCCCTTTTG	AAATTTTATG	TAACACTCAA	TTATATTTAT	GTATATGTAT
1170	SM4	551	GTCCCTTTTG	AAATTTTATG	TAACACTCAA	TTATATTTAT	GTATATGTAT
1171	SM5	551	GTCCCTTTTG	AAATTTTATG	TAACACTCAA	TTATATTTAT	GTATATGTAT
1172	SM6	551	GTCCCTTTTG	AAATTTTATG	TAACACTCAA	TTATATTTAT	GTATATGTAT
1173	SM7	551	GTCCCTTTTG	AAATTTTATG	TAACACTCAA	TTATATTTAT	GTATATGTAT
1174	SM8	551	GTCCCTTTTG	AAATTTTATG	TAACACTCAA	TTATATTTAT	GTATATGTAT
1175	SM9	551	tTaCaTgTTG	AAATTTTATG	TAACACTCAA	TTATATTTAT	GTATATGTAT
1176	SM10	551	tTaCaTgTgG	cAcTgTgAgG	gAcCcCgCcA	gTcTcTgTcT	tTcTcTtTcT
1177							
1178							
1179	yBE0.6	601	GCTCAAAATC	ACCTGCCAAT	AAC	CCTGCAG	G
1180	SM1	601	GCTCAAAATC	ACCTGCCAAT	AAC	CCTGCAG	G
1181	SM2	601	GCTCAAAATC	ACCTGCCAAT	AAC	CCTGCAG	G
1182	SM3	601	GCTCAAAATC	ACCTGCCAAT	AAC	CCTGCAG	G
1183	SM4	601	GCTCAAAATC	ACCTGCCAAT	AAC	CCTGCAG	G
1184	SM5	601	GCTCAAAATC	ACCTGCCAAT	AAC	CCTGCAG	G
1185	SM6	601	GCTCAAAATC	ACCTGCCAAT	AAC	CCTGCAG	G
1186	SM7	601	GCTCAAAATC	ACCTGCCAAT	AAC	CCTGCAG	G
1187	SM8	601	GCTCAAAATC	ACCTGCCAAT	AAC	CCTGCAG	G
1188	SM9	601	GCTCAAAATC	ACCTGCCAAT	AAC	CCTGCAG	G
1189	SM10	601	tCgCcAcAgC	cCaTtCaAcT	cAa	CCTGCAG	G
1190			<b>BE3.5 Rvs</b>				
1191							

**Figure S1. Sequence alignment of the yBE0.6 with scanning mutant versions.**

Purple background with white letters indicate the Ascl (GGCGCGCC) and SbfI (CCTGCAGG) restriction enzymes sites that were appended to primers for cloning CRE versions into matching sites in the S3aG reporter transgene vector. Maroon background and white letters indicate sequences that comprise scanning mutations. The lower case nucleotide letters indicate the non-complementary transversions. The yellow background with black letters indicates the Abd-B sites identified in Jeong et al. (2006) which were not mutated in this study. The blue background with white letters indicate the Bab-bound sequences identified in this study. The gray background with bolded black letters indicate the regions to which the BE2.5 Fwd and BE3.5 Rvs primers (reverse complement of that highlighted) were designed to initially amplify the wild type CRE sequence.

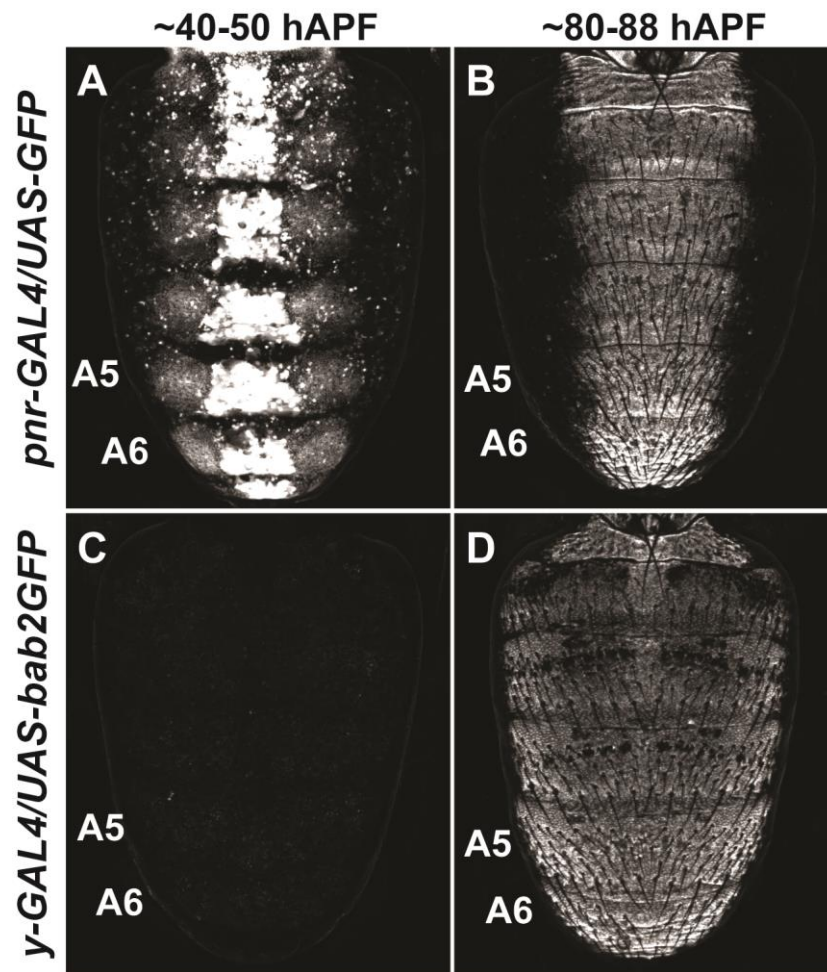


**Figure S2. Scanning mutagenesis across the entire yBE0.6 CRE identifies sequences that normally function to repress CRE activity in the female abdomen.** (A) Name and location of yBE0.6 scanning mutations. Scanning mutations for which CRE activity in the female abdomen was not noticeably altered are indicated as gray rectangle and those for which ectopic activity occurred are shown as light blue rectangles. The two vertical yellow lines on the illustration of the wild type CRE indicate the position of previously identified Abd-B binding sites that were not mutated in this study. (B-M) The EGFP reporter gene expression pattern in the female abdomen at ~85 hours after puparium formation driven by the non-mutant (yBE0.6) and scan mutant CRE sequences. (M) Scanning mutations 4 and 10 were combined together. Light blue arrowheads indicate abdomen segments with conspicuous ectopic EGFP expression.



ORF transgene	ectopic expressing		control	
	Observed	Expected	Observed	Expected
UAS-bab1 D. mel.	0	155.5	311	155.5
UAS-bab2 D. mel.	11	74.5	138	74.5
UAS-bab1 D. wil.	4	144.5	285	144.5
UAS-bab2 D. moj.	39	164.0	289	164.0
UAS-bab A. gam.	49	85.5	122	85.5

**Figure S3. Lethality from ectopic expression of orthologous *bab* open reading frame transgenes in the *pnr* pattern.** UAS-*bab* open reading frame transgenes are located in the attP40 site on the *D. melanogaster* 2<sup>nd</sup> chromosome. Male flies were crossed to females heterozygous for the 3<sup>rd</sup> chromosome where the GAL4 gene is inserted into the *pnr* locus and the TM3 balancer. Fewer offspring were obtained that possessed an ectopic *bab* expressing genotype than expected by chance, indicating lethality due to *bab* expression in the spatial and temporal pattern of the *pnr* gene.



**Figure S4. The temporal and spatial domains of activity for GAL4 drivers in *D. melanogaster* pupa.** (A) Dorsal midline expression of the UAS-GFP transgene under the control of the *pnr*-GAL4 driver at ~40-50 hours after puparium formation (hAPF). (B) Dorsal midline expression of the UAS-GFP transgene under the control of the *pnr*-GAL4 driver at ~80-88 hAPF. (C) The UAS-GFP transgene is not expressed at ~40-50 hAPF when under the regulatory control of the *y*-GAL4 driver. (D) The pan-abdomen expression of the UAS-GFP transgene under the control of the *y*-GAL4 driver at ~80-88 hAPF. The highest level expression occurs in the A5 and A6 segments due to the activity of the body element CRE which is included in the *y*-GAL4 transgene. All specimens shown are males.

# 1271 ***D. melanogaster bab1* ORF**

1272 gaattcaacttaaaaaaaaaaaatcaaaatggcgctcggcgcaggcggagacgaatgtcggc  
1273 M A S A Q A E T N V G  
1274 ttggcgctccgaacaggaccagtggtcagaggcagcgcaaaggacgggatcgggcgcc  
1275 L A S E Q G P V A Q R Q R K G T G S G A  
1276 gattcgcccaagagtaacagaagctcgccactcagcaggaggagaagcgtatcaaaagc  
1277 D S P K S N R S S P T Q Q E E K R I K S  
1278 gaggatcgcaacttaccactggcggggccaaggacgaggacaaggagagtcagggtcat  
1279 E D R T S P T G G A K D E D K E S Q G H  
1280 gctgtagccggagggggaggatcttcgcccgtcagttcgccacagggcaggagttcttcg  
1281 A V A G G G G S S P V S S P Q G R S S S  
1282 gtagcctcgcccagttccagctcccagcaattctgctgctggaacaactatcagacg  
1283 V A S P S S S S Q Q F C L R W N N Y Q T  
1284 aacctgaccaccatctttgaccagctgctccagaacgagtgcttcgtggacgtgaccttg  
1285 N L T T I F D Q L L Q N E C F V D V T L  
1286 gcatgcatggtcggtccatgaaggcccacaagatggctcctgtccgctgctcgccctac  
1287 A C D G R S M K A H K M V L S A C S P Y  
1288 ttccaaacacttctggcggaacgccttgccagcatccattgtgatcatgcgggacgta  
1289 F Q T L L A E T P C Q H P I V I M R D V  
1290 aattggctcgatctcaaggccattgtggagttcatgtatcgcggcgagatcaacgtgagc  
1291 N W S D L K A I V E F M Y R G E I N V S  
1292 caggaccagataggtcctctgctcaggatagctgagatgttgaaagtgcgtggtctggcg  
1293 Q D Q I G P L L R I A E M L K V R G L A  
1294 gatgtgaccatgatggaggcgccacggcagcagcggtgcgcttcgtcgagagagaatg  
1295 D V T H M E A G A T A A A S S E R M  
1296 ccctcctcgcccaaggagacacttcaacttcagaactgaacacgacaggggaacgggag  
1297 P S S P K E S T S T S R T E H D R E R E  
1298 gccgaggagctgctggccttcatgcagcccagaagaagctacgcacttcggactgggac  
1299 A E E L L A F M Q P E K K L R T S D W D  
1300 cccgctgagctgaggctctccccactggagcggcagcagggcaggaatgtaagaaagcgc  
1301 P A E L R L S P L E R Q Q G R N V R K R  
1302 cgggtggccatcgggcgacacaatatccaatccaccgcaccaccagtcactgagcagc  
1303 R W P S A D T I F N P P A P P S P L S S  
1304 ctgattgcgggccgaaggatggagctggagcaaaaggaaagagagagacagagggactgt  
1305 L I A A E R M E L E Q K E R E R Q R D C  
1306 tcgctgatgacacccccacccaaaccaccaatgagcagtggtccacagtgaggagccacg  
1307 S L M T P P P K P P M S S G S T V G A T  
1308 aggcgcctggagaccgccatccacgccttgacatgccatcgccggtgccacgccagga  
1309 R R L E T A I H A L D M P S P A A T P G  
1310 cctctgtcccgatcgctcgagacctcactcgagagccccagcagcagcagggcacagcag  
1311 P L S R S S R P H S Q S P Q Q Q A Q Q  
1312 cagggtcagcttctttgcccctgcccctgcacccaccatcacgcacacccgccccca  
1313 Q G Q L P L P L P L P L H P H H A S P A P  
1314 catccctcccagaccgcggatcagcccaccaccggcatcgctgctggagattccggt  
1315 H P S Q T A G S A H H P A S P A G D S R  
1316 tttcccctcgggccagcagccgcatggcgctgccagggaactgagtgccctgggacca  
1317 F P L G P A A A M A A A R E L S G L G P  
1318 ggtccgtccgcccagccacgccttccgcctccaccgcgcaccaccatggcggtggtgga  
1319 G P S A E P R L P P P P P H H H G G G G  
1320 gtggggcgggggggagttggaggaggaggtgcaggcggagtggttcaggcgggggatcc  
1321 V G G G G V G G G G A G G V G S G G G S  
1322 tcgctcgccgatgacttgagatcaagccagggatcgccgagatgatccgagaggaagaa  
1323 S L A D D L E I K P G I A E M I R E E E  
1324 agggccaaaatgatggagaactcgacgcctggatggcgccaccggatcaacgctggca  
1325 R A K M M E N S H A W M G A T G S T L A  
1326 gcagacagctaccagtaccagctgcagtcctatgtggcaaaagtgctggaacaccaaccag



1383 gctgtagccggagggggaggatcttcgcccgtcagttcgccacagggcaggagttcttcg  
 1384 A V A G G G G S S P V S S P Q G R S S S  
 1385 gtagcctcgcccagttccagctcccagcaattctgcctgcgctggaacaactatcagacg  
 1386 V A S P S S S S Q Q F C L R W N N Y Q T  
 1387 aacctgaccaccatctttgaccagctgctccagaacgagtgcttcgtggacgtgaccttg  
 1388 N L T T I F D Q L L Q N E C F V D V T L  
 1389 gcatgcgatggctcgggtccatgaaggcccacaaagatggctcctgtccgctgctcgccctac  
 1390 A C D G R S M K A H K M V L S A C S P Y  
 1391 ttccaaacacttctggccgaaacgccctgccagcatcccattgtgatcatgcgggacgta  
 1392 F Q T L L A E T P C Q H P I V I M R D V  
 1393 aattggctcggatctcaaggccattgtggagttcatgtatcgcgccgagatcaacgtgagc  
 1394 N W S S D L K A I V E F M Y R G E I N V S  
 1395 caggacagataggctcctctgctcaggatgagtgatgttgaaagtgcgtgggtctggcg  
 1396 Q D Q I G P L L R I A E M L K V R G L A  
 1397 gatgtgaccatattggaggcggccacggcagcagcggtgcgcttcgtcggagagaatg  
 1398 D V T H M E A A T A A A A A S S E R M  
 1399 ccctcctcgcccaaggagagcacttcaacttcagaactgaacacgacaggaacgggag  
 1400 P S S P K E S T S T S R T E H D R E R E  
 1401 gccgaggagctgctggccttcattgcagccccgagaagaagctacgcacttcggactgggac  
 1402 A E E L L A F M Q P E K K L R T S D W D  
 1403 cccgctgagctgaggctctccccactggagcggcagcagggcaggaatgtaagaaagcgc  
 1404 P A E L R L S P L E R Q Q G R N V R K R  
 1405 cgggtggccatcgccggacacaatattcaatccaccgcaccacccagtcactgagcagc  
 1406 R W P S A D T I F N P P A P P S P L S S  
 1407 ctgattgcggccgaaaggatggagctggagcaaaaggaaagagagagacagagggactgt  
 1408 L I A A E R M E L E Q K E R E R Q R D C  
 1409 tcgctgatgacacccccacccaaaccaccaatgagcagtggtccacagtgggagccacg  
 1410 S L M T P P P K P P M S S G S T V G A T  
 1411 aggcgcctggagaccgccatccacgccttgacatgccatcgccggctgccacgccagga  
 1412 R R L E T A I H L D M P S P A A T P G  
 1413 cctctgtcccgcctcgtcgagacctcactgcagagccccagcagcagcagggcacagcag  
 1414 P L S R S S R P H S Q S P Q Q Q A Q Q  
 1415 cagggtcagcttctcttggccctgcccctgcatccgcaccatcacgcacacccgccccca  
 1416 Q G Q L P L P L P L H P H H H A S P A P  
 1417 catccctcccagaccgcggatcagcccaccacccggcatcgctgctggagattccgct  
 1418 H P S Q T A G S A H H P A S P A G D S R  
 1419 ttccccctcgcccagcagccgcatggcgcgtgccagggaaactgagtggcctgggacca  
 1420 F P L G P A A A M A A A R E L S G L G P  
 1421 ggtccgtccgcccagccacgccttccgcctccaccgccgcaccaccatggcggtgggtgga  
 1422 G P S A E P R L P P P P P H H H G G G G  
 1423 gtggggcgggggggagttggaggaggaggtgcaggcggagtggttcaggcggggggatcc  
 1424 V G G G G V G G G G A G G V G S G G G S  
 1425 tcgctcgccgatgacttgagatcaagccagggatcgccgagatgatccgagaggaagaa  
 1426 S L A D D L E I K P G I A E M I R E E E  
 1427 agggccaaaatgatggagaactcgacgcctggatggcgccaccggatcaacgctggca  
 1428 R A K M M E N S H A W M G A T G S T L A  
 1429 gcagacagctaccagtaccagctgcagtcctatgtggcaaaagtgtggaacaccaaccag  
 1430 A D S Y Q Y Q L Q S M W Q K C W N T N Q  
 1431 aatctgatgcatacatcgcttccgcgagcgaggtcctctgaagtgcgtggcgaccgag  
 1432 N L M H H M R F R E R G P L K S W R P E  
 1433 accatggcgaggGgCCctttagtggtctaaaggagggtctatcgctatctcaggccgcc  
 1434 T M A E G P F S V L K E G L S L S Q A A  
 1435 cgcaagtacgacatcccgatatccaacattcgtgctctatgcgaacaggggtgcacaatatg  
 1436 R K Y D I P Y P T F V L Y A N R V H N M  
 1437 ctgggaccatccattgacggcgggcccgatttgcggcccaaggggGAtggcGACccgcag  
 1438 L G P S I D G G P D L R P K G D G D P Q  
 1439 cgaatccttctgggcatctggcccgcagcagcattaaagggcgctcatcaagacgggtggtc



1496 gagaccaagatcaaaaccaatccagagacaaaaccgcccagggcgcaaaatagttcctccc  
 1497 E T K I K T N P E T K P P R R K I V P P  
 1498 agcggcgaggggagcaggttctgcctgaggtggaacaactatcagtctaacctgaccaat  
 1499 S G E G Q Q F C L R W N N Y Q S N L T N  
 1500 gtcttttgacgaactccttcagagcgagtccttcgtggacgtgaccttgctcctgcaaggc  
 1501 V F D E L L Q S E S F V D V T L S C E G  
 1502 cactcgatcaaggcacacaagatggtgctatccgcctgctcaccctacttccaggccctg  
 1503 H S I K A H K M V L S A C S P Y F Q A L  
 1504 ttctacgacaatccctgccagcaccccatcatcatcatgcgggacgtcagctgggtccgac  
 1505 F Y D N P C Q H P I I I M R D V S W S D  
 1506 ctgaaggccctggtggagttcatgtacaaggggagatcaacgtctgccaggatcagata  
 1507 L K A L V E F F M Y K G E I N V C Q D Q I  
 1508 aaccctgctcaaagtggccgaaacctgaagatcaggggtctggcgagggtcagtgcg  
 1509 N P L L K V A E T L K I R G L A E V S A  
 1510 ggcaggggagggaggcgccctccgcacttcccatgtccgccttcgacgatgaggacgag  
 1511 G R G E G G A S A L P M S A F D D E D E  
 1512 gaggaggaactggcctcgccactgcaattctgcagcaggacggtgatgccgatcccgat  
 1513 E E E L A S A T A I L Q Q D G D A D P D  
 1514 gaggagatgaaggccaagaggcccagactgctgcccagggaggtcttggaacttgaatcag  
 1515 E E M K A K R P R L L P E G V L D L N Q  
 1516 cgacaaaggaagcgggtccagggtggcagctacgccactccaagtccatcccttcagggc  
 1517 R Q R K R S R D G S Y A T P S P S L Q G  
 1518 ggagagtccgagatctcggagaggggtcatccggcactccgggacagagccagagccaa  
 1519 G E S E I S E R G S S G T P G Q S Q S Q  
 1520 cccctggccatgaccacctccaccattgtgcgcaatcccttcgcctccccaatcctcag  
 1521 P L A M T T S T I V R N P F A S P N P Q  
 1522 accttgaggggcaggaacagcgccatgaatgcagtagcaaaccagaggaaatcaccagca  
 1523 T L E G R N S A M N A V A N Q R K S P A  
 1524 ccaacagcgacaggtcacagcaatgggaacagcgggcgccgatgcactccccacccggg  
 1525 P T A T G H S N G N S G A A M H S P P G  
 1526 ggcgtggcggtccagtcgcgccttcgcgcacacatggcgccatcgtgccgcccaccccc  
 1527 G V A V Q S A L P P H M A A I V P P P P  
 1528 tccgcatgcaccatcatgcccagcaactggcgcccagcaccagctggcccactcgcac  
 1529 S A M H H H A Q Q L A A Q H Q L A H S H  
 1530 gccatggccagcgcccttgagcgcgcagcgcggagctggcgagcgaggagcgggcgga  
 1531 A M A S A L A A A A A G A G A A G A G G  
 1532 gcaggatctggcagtggtatcgggcgccagtgctccgactggaggaacaggagtgggcgga  
 1533 A G S G S G S G A S A P T G G T G V A G  
 1534 agtggagccggcgcggtgggatcccatcacgatgacatggagatcaagccagaaatc  
 1535 S G A G A A V G S H H D D M E I K P E I  
 1536 gccgagatgatacggaagaagagagggccaagatgatcgagagtggaggccacggtgga  
 1537 A E M I R E E E R A K M I E S G G H G G  
 1538 tggatgggagcggcagctgcggaactggagcagcttctgtggcggcagatagctaccag  
 1539 W M G A A A A A T G A A S V A A D S Y Q  
 1540 taccagctacagtccatgtggcagaagtgtggaacaccaatcagcagaacctggtgcag  
 1541 Y Q L Q S M W Q K C W N T N Q Q N L V Q  
 1542 cagctcagattccgcgagcgcgccattgaagtccctggcgacccgaggccatggccgag  
 1543 Q L R F R E R G P L K S W R P E A M A E  
 1544 gccattttcagtgctcctgaaggaggggtctcctgtcacaggctgcccgcaagttcgac  
 1545 A I F S V L K E G L S L S Q A A R K F D  
 1546 ataccctatcccaccttcgtcctgtacgccaatcggtgcacaacatgctgggaccctcg  
 1547 I P Y P T F V L Y A N R V H N M L G P S  
 1548 ctggatggcgagctgatccgcggccaaaggcacgcggtcgtccccagaggatcctgctg  
 1549 L D G G A D P R P K A R G R P Q R I L L  
 1550 ggcagtggtggcgaggagctcatccgtagcgtcattaaggccgtggtgttcgggactat  
 1551 G M W P E E L I R S V I K A V V F R D Y  
 1552 cgcgagattaaggaggacatgagcgcccatcagtagccaatggacaggggtcatggtacc

1553 R E I K E D M S A H Q Y A N G Q G H G T  
1554 tatatcggaggaggaaccaccacgaatggctaccacagtgtctgccgcagccaagctggcg  
1555 Y I G G G T T T N G Y H S A A A A K L A  
1556 gctcagaacgctgcactggctccgcccgcagcaggaagtccgctgagctccatgacggaa  
1557 A Q N A A L A P P D A G S P L S S M T E  
1558 acccttcgccgcagatcctctcgcagcagcagcaacatcagcagcaccaccagcagcag  
1559 T L R R Q I L S Q Q Q Q H Q Q H H Q Q Q  
1560 gcacaccatcagcaacagccctcgcaccaccagcaacagtcgccccacgcccagtcctatg  
1561 A H H Q Q Q P S H H Q Q Q S P H A Q S M  
1562 aacatgtacaagtccccggcctatctgcagcagatccgagatcgaagatcaagtatccgca  
1563 N M Y K S P A Y L Q R S E I E D Q V S A  
1564 gcggcgccgctggcagcggcgccgccaagcaccagcagcagcaggggtgagcgaaggggt  
1565 A A A V A A A A A K H Q Q Q G E R R G  
1566 tcggagaacctgcccgacctcagtgccctgggctgatgggtctgccggcctgaatgtg  
1567 S E N L P D L S A L G L M G L P G L N V  
1568 atgccctcacggggatcgggtggaggaagtgggtggcgcagcgcgaatagtgccgcctcc  
1569 M P S R G S G G G S G G A A P N S A A S  
1570 tatgccgcgagttatcccgcgaaaggggaacgcgacatcgggagcgcgaaagggagcgggag  
1571 Y A R E L S R E R E R D R E R E R E R E  
1572 ctgtcccgccagtatggcagccagtcgcggggatcgagctccggttcggaagcgccaag  
1573 L S R Q Y G S Q S R G S S S G S G S A K  
1574 tccctgaccgccagccaaagaccaggagccgcctcgcggtactccgcccgcactatgcc  
1575 S L T A S Q R P G A A S P Y S A A H Y A  
1576 aaacatcaggcgagtgccctacaacaagaggtttctcgagagcctgcccgccggcattgac  
1577 K H Q A S A Y N K R F L E S L P A G I D  
1578 ttggaggccttcgccaacggactgctccagaagtccgtgaacaagagtcgcgccttcgag  
1579 L E A F A N G L L Q K S V N K S P R F E  
1580 gacttcttcccgggaccggccaggacatgagtgaactgtttgccaatccggacgcgagt  
1581 D F F P G P G Q D M S E L F A N P D A S  
1582 gcagtgccgcggcgccgacctacgcgcctcctggcgccatccgcgaatcgccctctgatg  
1583 A A A A A A A Y A A P P G A I R E S P L M  
1584 aagatcaagctggagcagcagcatgccaccgaactgccgcacgaggattgataggcgcc  
1585 K I K L E Q Q H A T E L P H E D - - - -  
1586 Gc

# **A. gambiae bab ORF**

1590 gaattcaacttaaaaaaaaaaatacaaaatgggcaagccaatcccgaacccccctgctgggc  
1591 M G K P I P N P L L G  
1592 ctggactccacaccaagcgatacacccccgcgcagcgcacctccgtgagccaccatcg  
1593 L D S T P S D T P P P S A T S V S H P S  
1594 cccgccagttcgcaccatgatcccaacgatccaaatgccccccgcgcgatcccggtggat  
1595 P A S S H H D P N D P N A P P R D P V D  
1596 cgcagtgccacaggaacccccgggccccagcgatcacccaacaggcggacacctgggccac  
1597 R S G T G T P G P S D H P T G G H L G H  
1598 catcagccaccgtcctcgtcgagtagctcctcgcagctcgagttccagctcgtcgacaagc  
1599 H Q P P S S S S S S S S S S S S S S T S  
1600 tcctcgtgtcctcgtgagtgctgaagcgtcctcggaggagccccctgacgaccgccaag  
1601 S S L S S L S L K R S L E E P L T T A K  
1602 ccattcgccgcctcgcagccccctgaccatggaccaccatcatcagcataaggccgcccgc  
1603 P S P P C S P L T M D H H H Q H K A A R  
1604 cagtcgcgcgcgcctcgcgcgcggacgcagtagcgcagcagcagggcctcgccaagtgcc  
1605 Q S R A A S P A G R S T Q Q Q A S P S A  
1606 cccggcaccggcggtatcgagtagcggaggcgggaggaggccagcagttctgcctgcgc  
1607 P G T G G S S S G G G G G G Q Q F C L R  
1608 tgggaataactaccagaccaatctgaccagcgtgttcgaccagctgctgcagagcagtgctg

1609	W N N Y Q T N L T S V F D Q L L Q S E S
1610	ttcgtggatgtgacctggcgctgcgatggccagagcatgaaggcccataaatggtgctg
1611	F V D V T L A C D G Q S M K A H K M V L
1612	agcgctgctcgccgtacttccagacgctgtttttcgataaccgctgccagcatccatt
1613	S A C S P Y F Q T L F F D N P C Q H P I
1614	gtgatcatgcgcgacgtgtcgtgggccgagctgaaggccatcgtggagtcatgtacaag
1615	V I M R D V S W A E L K A I V E F M Y K
1616	ggcgagatcaatgtgtcgcaggatcagatcggccccctgctgaaggtggccgagatgctg
1617	G E I N V S Q D Q I G P L L K V A E M L
1618	aagatccgcgccgtggccgatgtgagcggagacgccggagagccaacaggaaagccgcgc
1619	K I R G L A D V S G G D A G E P T G S R A
1620	gagcgcgaggcccggaagccgcggccccgaggagctggatcgcgaggagcatggcaag
1621	E R E A A G S R G P E E L D R E E H G K
1622	ctgctgaacccccctggccatcgtgggatcgagcctgctggccaatggagccgccagcgcc
1623	L L N P L A I V G S S L L A N G A A S A
1624	gccatggccggaggcaacggcagtaacagcacccgccacaagcggtccgcgcgctgcag
1625	A M A G G N G S N S T A T S G S A A V Q
1626	gccgcgcgcgcgcgcgcgcgcgcgccaagaagcagcgcgcgcggacgcgatcgcgacacaacc
1627	A A A A A A A A K K Q R A G R D R D T T
1628	aaggagcaccgcatggatgccgcctgtcggagtttgccgcgcacctgagccgcgcgat
1629	K E H R M D A R L S E F A R D L S R A D
1630	ccccacatctcgagccgcgatatcagcagtggtggccgcgcgcgcgcgcgcgcgcgcgcgc
1631	P H I S S R D I S S V A A A A A A A A A
1632	gccgcgcgcggcctggccgtgggagagtggccccctgggagccgcgcgcgcctggaggccgc
1633	A A A G L A V G E W P L G A A G L E A A
1634	gccgcgcgcgcgcgcgcgcgcgcgccaagtcgccccgcaagcgcgcgcgtggccctcg
1635	A A A A V Q A S T P K S A R K R R W P S
1636	ggagagcgctcgagcattggatcgccagccgacagcacccccggaccagctggaggtgcca
1637	G E R S S I G S P A D S T P D Q L E V P
1638	tcgccgatcccacccacaccgagtagcctggcccagtcgagcggaggaggcgaggcggc
1639	S P I P P T P S S L A Q S S G G G G G G
1640	ggcgggcgaggcgaggcggcacaggaagtggcggcgcgcgcgcgcgcgcgcgcgcgcgcgc
1641	G G G G G G G T G S G G G G G G S S N P
1642	ctggcctcctttccgctgccaccgcgcctggacaccgcgcgcctatggccatgagtagcctg
1643	L A S F P L P P A L D T A A M A M S S L
1644	tccagttcgatcgccaatcacccagacgacatggagatcaagccgggaattgccgagatg
1645	S S S I A N H P D D M E I K P G I A E M
1646	atccgcgaggaggagcgcagcatgtggcagaagtgtggaacagccagaacctgatccac
1647	I R E E E C R S M W Q K C W N S Q N L I H
1648	catrtgcgttttcgcgagc
1649	H L R F R E R G P L K S W R P E T M A E
1650	gccatcttcagcgtgctgaaggaggactgctgctgagccaggccgcgcgcgcgcgcgcgcgcgc
1651	A I F S V L K E G L S L S Q A A R K Y D
1652	atcccatatcccaccttcgtgctgtacgccaaccgcgtgcataaacatgctgggccccagc
1653	I P Y P T F V L Y A N R V H N M L G P S
1654	atcgatggaggcacccgacctgcgcaccaagggccgcgcgcgcgcgcgcgcgcgcgcgcgcgc
1655	I D G G T D L R P K G R G R P Q R I L L
1656	ggcatctggcccgacgatcacatcaagggagtgatcaagtccgtggtgttcgcgcgatgcc
1657	G I W P D D H I K G V I K S V V F R D A
1658	aaggacatgaaggaggagccgatgatgtatggacgccacagtcogttcccccttcaggat
1659	K D M K E E P M M Y G R H S P F P F Q D
1660	aaccgcgtgagctacggaccaaccgccccaaaatggccagctgccctcggtggccacaggc
1661	N P L S Y G P T A P N G Q L P S V A T G
1662	accaacgtgcccgatggcatgagccaggacgcacctgaccgcgcgcacagtggccgcgcgtg
1663	T N V P D G G M S Q D A L T A A T V A A V
1664	cgcagcagcatgtgcaacatgggtggccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc
1665	R O O M C N M V A A A O H H P D A A N L

1666 gtggccgcccgcggatttaacctgccatcccactgcggcacccccccgaatctgtcgatg  
1667 V A A A G F N L P S H C G T P P N L S M  
1668 caccagccgcccgcggccgcccgcggccgcccgcggccgcccgcgctcgaatgcctcgccgccc  
1669 H P A A A A A A A A A A A S N A S A A  
1670 ggaggcccgtccggaggaggaggcggaggaggctcgagcggcgccattccgctgccgaag  
1671 G G P S G G G G G G G S S G A I P L P K  
1672 atgggatcgccagccgtgccgtccacgggccacggaacaataacggcggaagtggcgcc  
1673 M G S P A V P S T G H G N N N G G S G A  
1674 ggcattcagatgccacgcctgggaagccccgcggatcgagcggcctggccaaggagcat  
1675 G I Q M P R L G S P A G S S G L A K E H  
1676 gagctgcagcatcacggcggaggaggaggaggcggcggcctgggaggcggatcgggc  
1677 E L Q H H G G G G G G G G G G G G G S G  
1678 ggaggaatgagcgcgccacccccccggcgcccgatcgcgccatgaccgcccgtcc  
1679 G G M S R A T P P G A R D R A M T A R S  
1680 aatctggccgcccggagagaccggacgcagcagttcgagcggcgctcgatccaccgctcc  
1681 N L A A G E T G R S S S S A G S I H R S  
1682 agcccgagctcctcgccggcgtcgagtctgaatcaccagcatcccgcccacctgtcccac  
1683 S P S S S A G S S L N H Q H P A H L S H  
1684 ccacaccatcagcagcagcatcaccatcagcatcatcaccagccgcaccatggacatgcc  
1685 P H H Q Q Q H H H Q H H H Q P H H G H A  
1686 catcacctgccacatcacaacccccctgagtcattctggtgggatcgggaggagccagcggc  
1687 H H L P H H N P L S H L V G S G G A S G  
1688 gccctgagcatcaccaagctgggctcccccggaagtgccacgatctgcgcatctcgaat  
1689 A L S I T K L G S P G S A H D L R I S N  
1690 agtcccgcagagagtccactggccagcccaattggactggccatggagccagccgtgaat  
1691 S P D E S P L A S P I G L A M E P A V N  
1692 ctggccctggggcgccggagggaacccagccccggccccgaggatgtgcgcctgcatgtgcc  
1693 L A L G A G G T Q P G P E D V R L H V P  
1694 ccgccatacggctcgaagccgcggcggcggcggaggagccccagcaccggatacacg  
1695 P P Y G A S K P P S R G G A P S T G Y T  
1696 agcaactcgagtcggccccgcccagagcatctgttccaggatcaggacattgccgcctg  
1697 S N S S P P R P E H L F Q D Q D I A A L  
1698 gtggccaccacgcgcgcgcctgccccccagccgcgtgcccgattacaaggacaccgcc  
1699 V A T T R A A C P P S R V P D Y K D T A  
1700 gtgcgcccgcagaccagcatcaaggtggagccctgacagagtgcgcgggagactaatag  
1701 V R P T A S I K V E P L T E C R G D - -  
1702 gcggccgc  
1703 - -  
1704

#### ***D. mojavensis bab2* ORF**

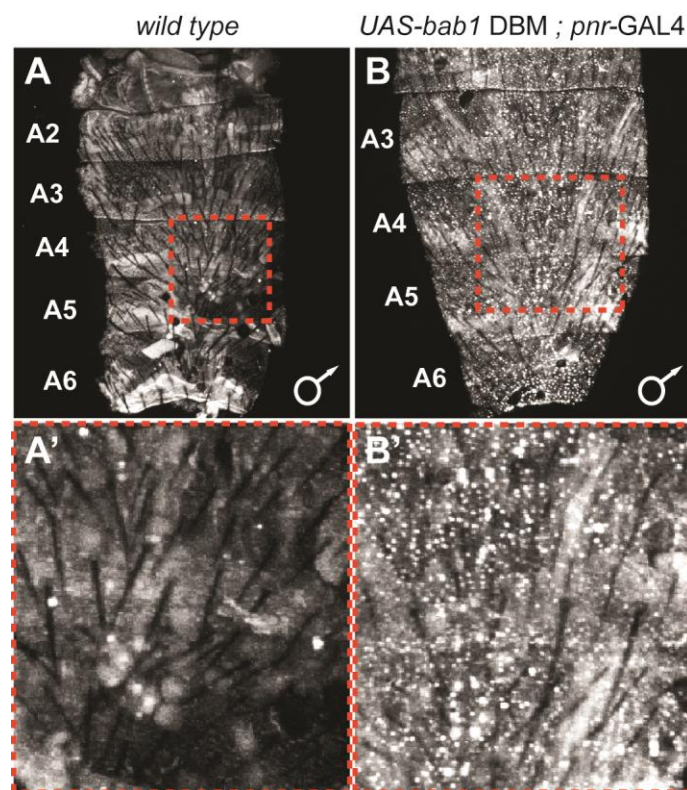
1706 gaattcaacttaaaaaaaaaaaaatcaaaatggacatgacaaaggacattatggacttcgag  
1707 M D M T K D I M D F E  
1708 cgcaagagccctggatagttcgtgcccagagcagttcgagccgtcggattacacaatgggtg  
1709 R K S L D S S C G E Q F E P S D Y T M V  
1710 aatgccgagctggccaagcaggccgcccagaccgcccaggccgtggatcaggtggagctg  
1711 N A E L A K Q A A Q T A Q A V D Q V E L  
1712 gacctgccactggagctggccaagaaggaggagccggaggcccgagcccgagcccatgcag  
1713 D L P L E L A K K E E P E A Q P E P M Q  
1714 cagctgaaggaggagaatcgcgccgtggccgcccagagaagccccgcatgctgaatgagcag  
1715 Q L K E E N R A V A A E K P A M L N E Q  
1716 gccctgacccccccaccgcgccccctgaccagcagcaggtggtgggcccactccgagcca  
1717 A L T P P P R P L T S S E V V G H S E P  
1718 tcggaccggagctgcagattcagctgaccgccaagaagtcgcgcagcctgccggtgtcg  
1719 S D P E L Q I Q L T A K K S R S L P V S  
1720 ccacagccactgggtggcccataatctggccgcatcggaactgtttgagttcggaagacg  
1721 P Q P L V A H N L A A I G L F E F G K T

1722 gtggagacacccgagctgaagcccaagatgaatcacaagctgctgcccccggtgaacgtg  
 1723 V E T P E L K P K M N H K L L P P V N V  
 1724 ggcgtggccccacgcaaggtggccccagcgccggcggaggcgacaatcagcagttctgc  
 1725 G V A P R K V A P S A G G G D N Q Q F C  
 1726 ctgcgctggaataactaccagagcaacctgaccaacgtgttcgacgagctgctgcagaac  
 1727 L R W N N Y Q S N L T N V F D E L L Q N  
 1728 gagtcctttgtggacgtgaccttgccctgcatggccagagcattaaggcccacaaagatg  
 1729 E S F V D V T L A C D G Q S I K A H K M  
 1730 gtgctgtcggcctgctccccctacttccaggccctgttctacgataacccgtgccagcac  
 1731 V L S A C S P Y F Q A L F Y D N P C Q H  
 1732 cccattatcatcatgcgcgatgtgaactgggtgcgacctgaaggccctgggtggagttcatg  
 1733 P I I I M R D V N W C D L K A L V E F M  
 1734 tacaaggagagatcaacgtgtgcccaggacagattaatcccctgctgaaggtggccgag  
 1735 Y K G E I N V C Q D Q I N P L L K V A E  
 1736 accctgaagattcgcggcctggccgaggtgggcgccctcgccaccgcccgcggcctgggc  
 1737 T L K I R G L A E V G A S S T A A G L G  
 1738 gccgccagcatgctgcccagcagcgcatgagcgtgtatgacgatgaggaggatgaggac  
 1739 A A S M L P E Q R M S V Y D D E E D E D  
 1740 gagctggccgcccgcggccgcccctgctgaacgatgaggatgaggatgagctgctgaagcca  
 1741 E L A A A A A L L N D E D E D E L L K P  
 1742 aagcgcgcccgcctgctggccaagctgcgcgcggccgagaccgcccctggatctgaaccag  
 1743 K R A R L L A K L R A A E T A L D L N Q  
 1744 cgccagcgcaagcgctcccgcgatggcagctacgccacccccctcgccactgcgagcgag  
 1745 R Q R K R S R D G S Y A T P S P L R S E  
 1746 tcgcccagattcgcagctgccactggccatgacgaccagcaccattgtgcgcaatcccttt  
 1747 S P S S Q L P L A M T T S T I V R N P F  
 1748 gccagccccaatccccagaccctgccagcctcgagtggagttcgtccaacagtaacagc  
 1749 A S P N P Q T L P A S S G S S S N S N S  
 1750 aataacagctcgtgcaacaactcgtccagcaacagttccagcaccgcccacagccgcccgc  
 1751 N N S S S T A T A A A  
 1752 cagccgaccgcccacaaactgcagcagctccagtagcgccggcggtgccaagcaacggaagt  
 1753 Q P T A T N C S S S S A G V P S N G S  
 1754 agctcggccgcctatcgagctccacccccaccgcccccccccccgctcgctccgcccatagc  
 1755 S S A A Y R S P P P P P P P S S A H S  
 1756 aatggatcgagcgccgcccggactgagctcgccacaggaaacaagagctccgcccgcgcc  
 1757 N G S S A A G L S S P T G N K S S A A A  
 1758 gccgcccagcagccagctgccacccccatatggccgcccgcgtggccgcccgcgcccat  
 1759 A A A Q S Q L P P H M A A A V A A A A H  
 1760 cagcctccgccaatgtgccgccccaccaccaggagccgcccgcctcgatgcaccatcac  
 1761 H A S A N V P P P P P G A A A S M H H H  
 1762 gccgcccgcggccgcccagcagctggccgcccagcaccagctggccacagccatgcc  
 1763 A A A A A A Q Q L A A Q H Q L A H S H A  
 1764 gccatggccagcgtgctgggcgcctcgctggccgcccgcggcggcggaggcgccgcccgc  
 1765 A M A S V L G A S L A A A A A G G A A A  
 1766 cccggctccgcccgcggcgccggaatgccccaaagctcgggtgggaggacaccatgacgat  
 1767 P G S A A G A G A G N A P S S V G G H H D D  
 1768 atggagatcaagcccagattgccgagatgattcgcgaggaggagcgccgaagatgatc  
 1769 M E I K P E I A E M I R E E E R A K M I  
 1770 gagaccagcggccatgcctggatgggcgccccagccacaggagcctcgggtggccgcccgc  
 1771 E T S G H A W M G A P A T G A S V A A D  
 1772 agctaccagtagcagctgcagagcatgtggcagaagtgctggaataaccaaccagcagaac  
 1773 S Y Q Y Q L Q S M W Q K C W N T N Q Q N  
 1774 ctgggtgcagcagctgcgcttttcgcgagcgccgaccactgaagtcgtggcgccccgaggcc  
 1775 L V Q Q L R F R E R G P L K S W R P E A  
 1776 atggccgaggccatcttttcgggtgctgaaggaggaggactgagctctgagccaggccgcccgc  
 1777 M A E A I F S V L K E G L S L S Q A A R  
 1778 aagtacgacatcccataccccacctttgtgctgtatgccaaccgctgcataatatgctg

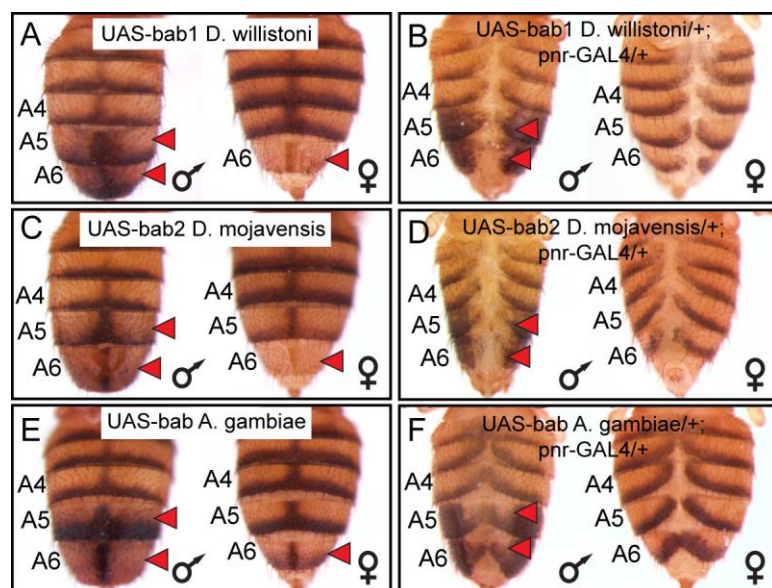




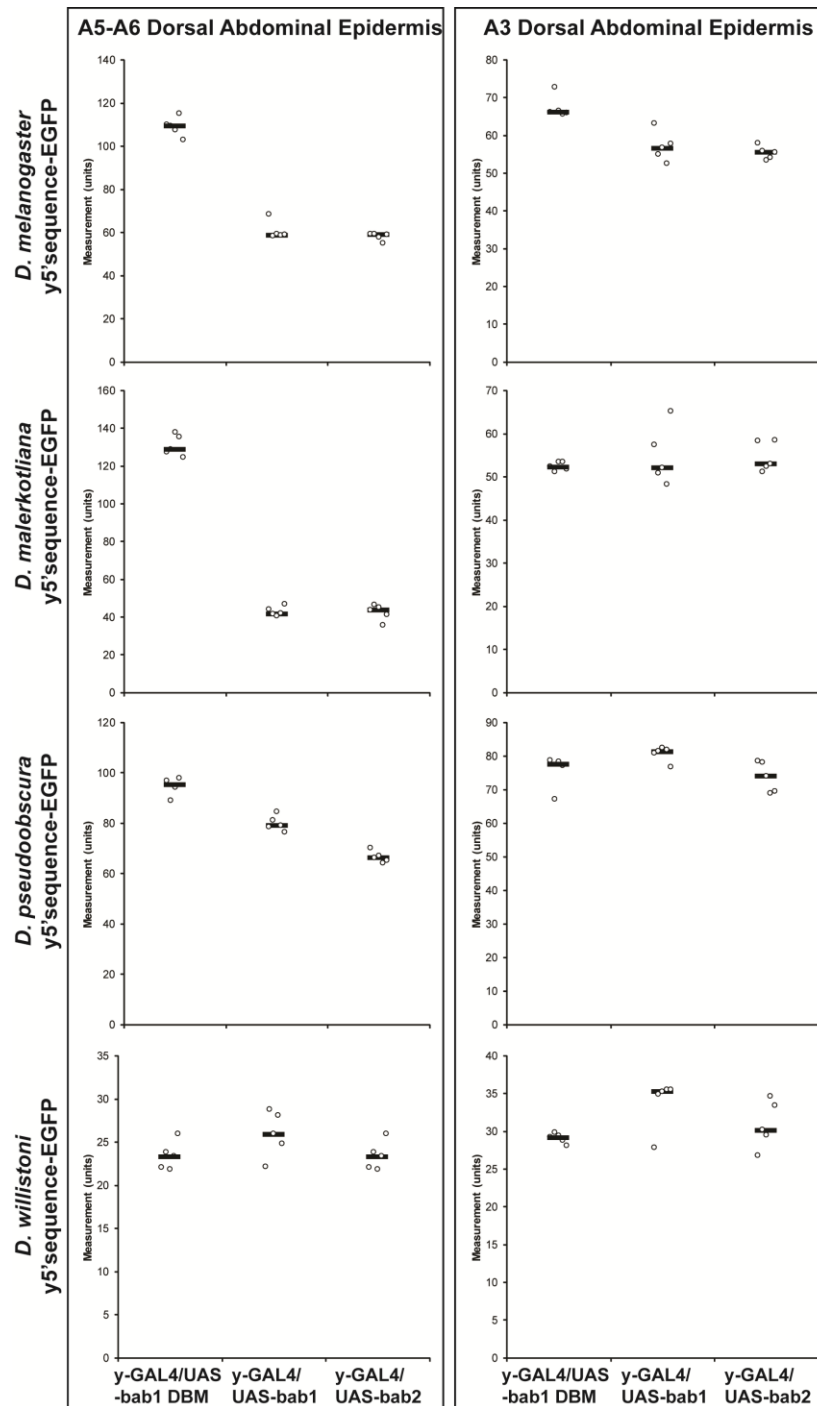




**Figure S6. Ectopic expression of the Bab1-DNA binding mutant protein.** (A) Little-to-no endogenous Bab1 protein can be detected in the dorsal abdominal epidermis of *D. melanogaster* male pupa. (B) In contrast, nuclear-localized expression of the Bab1 DNA-binding mutant protein can be observed when the UAS-transgene was ectopically expressed under the control of the *pnr*-GAL4 driver. (A' and B') Zoomed in view of the expression within the regions outlined by dashed red boxes in panels A and B. Samples shown are at a pupal developmental stage of ~88 hours after puparium formation.



**Figure S7. Orthologous Bab proteins are sufficient to suppress tergite pigmentation in *D. melanogaster*.** (A, C, and E) Leaky expression of transgenes from the attP40 transgene insertion site results in suppression of tergite pigmentation. (B, D, and F) Ectopic expression of protein coding sequences in the dorsal midline of male and female abdomens under control of the *pnr*-GAL4 driver causes a reduction in pigmentation and a split tergite phenotype. Red arrowheads indicate tergite regions with notably reduced tergite pigmentation.



**Figure S8. Orthologous regulatory regions 5' of the *yellow* gene differ in their responsiveness to *bab*.** Scatter plots of the pixel intensity statistics obtained for the EGFP reporter expression occurring in the dorsal abdominal epidermis of the A5 and A6 segments and the A3 segment. For each condition (a reporter transgene with expression driven by a *yellow* CRE and an ectopically expressed *bab* open reading frame), expression was measured for 5 replicate male specimens for which the mean is shown as a horizontal black bar. All specimens used were at the developmental stage of ~85 hours after puparium formation for growth at 25°C.

**Table S1. Design of Small Interfering RNA output for the *bab1* ORF.**

siRNA id	Position	SS Sequence (Passenger)	AS Sequence (Guide)	Corrected Score
1	560	GGAUAGCUGAGAUGUUGAAAG	UUCAACAUCUCAGCUAUCCUG	99.7
2	1442	CCGAUGACUUGGAGAUCAAGC	UUGAUCUCCAAGUCAUCGGCG	85.7
3	278	GGAACAACUAUCAGACGAACC	UUCGUCUGAUAGUUGUUCAG	97.6
4	162	GAGUCAAGGUCAUGCUGUAGC	UACAGCAUGACCUUGACUCUC	95.2
5	1483	CGAGAGGAAGAAAGGGUAAGU	UUACCCUUUCUUCCUCUCGGA	81.8
6	150	CGAGGACAAGGAGAGUCAAGG	UUGACUCUCCUUGUCCUCGUC	94.6
7	1473	CGAGAUGAUCCGAGAGGAAGA	UUCCUCUCGGAUCAUCUCGGC	78.9
8	219	GGGCAGGAGUUCUUCGGUAGC	UACCGAAGAACUCCUGCCCUG	89.5
9	359	GCGAUGGUCGGUCCAUGAAGG	UUCAUGGACCGACCAUCGCAU	88.2
10	664	CCCAAGGAGAGCACUUAACU	UUGAAGUGCUCUCCUUGGGCG	84.7
11	368	GGUCCAUGAAGGCCCAACAAGA	UUGUGGGCCUUAUGGACCGA	87.5

2008 **Table S2. Design of Small Interfering RNA output for the *bab2* ORF.**

siRNA _id	Position	SS Sequence (Passenger)	AS Sequence (Guide)	Corrected Score
6	16	GAUUGUGGACUUUGAAUAAA	UAUUUCAAAGUCCACAAUCUG	98.1
12	279	CGGAGCUGGUGAAGUCCAAGG	UUGGACUUCACCAGCUCCGUU	94.5
20	51	GCGAAAUCGAUCAGUUCGAGG	UCGAACUGAUCGAUUUCGCCG	94.4
18	155	AGAAAGUACUCACCCGAAAGG	UUUCGGGUGAGUACUUUCUGU	93.6
19	202	AAGUGAGGUGGUUGAUCAAU	UUGAUCAACCACCUCACUUGG	92.5
23	241	CGUUGGAGAAGUCAAGUCACC	UGACUUGACUUCUCCAACGCU	92.3
42	1	GGACAUGACCAAACAGAUUGU	AAUCUGUUUGGUCAUGUCCAU	91.7
43	14	CAGAUUGUGGACUUUGAAUA	UUUCAAGUCCACAAUCUGUU	91.6
45	63	AGUUCGAGGCGAGUGACUACA	UAGUCACUCGCCUCGAACUGA	91.4
40	154	CAGAAAGUACUCACCCGAAAG	UUCGGGUGAGUACUUUCUGUU	90.7
49	13	ACAGAUUGUGGACUUUGAAU	UUCAAAGUCCACAAUCUGUUU	90.7
28	306	CGAUGAACGACCAAGCUUUGA	AAAGCUUGGUCGUUCAUCGGA	90.6
46	140	CUAGAGGACCAGAACAGAAAG	UUCUGUUCUGGUCCUCUAGAU	90.4
13	625	GACCAAUGUCUUUGACGAACU	UUCGUCAAAGACAUUGGUCAG	90.3
55	12	AACAGAUUGUGGACUUUGAAA	UCAAAGUCCACAAUCUGUUUG	90.3
38	297	AGGCGAGUCCGAUGAACGACC	UCGUUCAUCGGACUCGCCUUG	89.8
27	443	CAGCCUCAACCAAAUCUUAAG	UAAGAUUUGGUUGAGGCUGUG	89.6
10	822	UGGUGGAGUUCAUGUACAAGG	UUGUACAUGAACUCCACCAGG	88.9
4	1099	GGACUUGAAUCAGCGACAAAG	UUGUCGCUGAUUCAAGUCCAA	88.8

2009

2010

2011

2012

2013

2014

2015

2016

2017

2018

**Table S3.** Oligonucleotides used to make Scan Mutant 4 region gel shift probes.

Probe	Sequence (5' to 3')	Oligo Name
SM4 Region 1	ATTCTTTAATTTGTATTTTAATATT	yBE 4i1 Top
	AATATTTAAAATACAAATTAAAGAAT	yBE 4i1 Bottom
SM4 Region 2	ATATTTTGAGAGGTTTTCTTATTTAAAGT	yBE 4i2 Top
	ACTTTAAATAAGGAAAACCTCTCAAAATAT	yBE 4i2 Bottom
SM4 Region 3	AAAGTGTAGATTATTGAGGATTAAT	yBE 4i3 Top
	ATTAATCCTCAATAATCTACACTTT	yBE 4i3 Bottom
SM4 Region 3 Scan Mutant	cAcGgGgAtAgTcTgGcGtAgTcAg	y4i3 T Scrm
	cTgAcTaCgCcAgAcTaTcCcCgTg	y4i3 B Scrm
Region 3 TA>GA	AAAGTGgAGATgATTGAGGATgAAT	yBE 4i3 TA>GA Top
	ATTcATCCTCAATcATCTcCACTTT	yBE 4i3 TA>GA Bottom
SM4 Region 3 sub1	gggCgggCgATTATTGAGGATTAAT	y4i3 sub1 T
	ATTAATCCTCAATAATcGgggGccc	y4i3 sub1 B
SM4 Region 3 sub2	AAAGTgggCgggCgTGAGGATTAAT	y4i3 sub2 T
	ATTAATCCTCAcGcccGcccACTTT	y4i3 sub2 B
SM4 Region 3 sub3	AAAGTGTAGAgggCgggCgATTAAT	y4i3 sub3 T
	ATTAATcGcccGcccTCTACACTTT	y4i3 sub3 B
SM4 Region 3 sub4	AAAGTGTAGATTATTGgggCgggCg	y4i3 sub4 T
	cGcccGcccCAATAATCTACACTTT	y4i3 sub4 B

**Table S4.** Oligonucleotides used to make Scan Mutant 10 region gel shift probes.

Probe	Sequence (5' to 3')	Oligo Name
SM10 Region 1	TCGTCCCTTTTGAATTTTATGTAACACTC	yBE 10i1 Top
	GAGTGTTACATAAAATTTCAAAGGGACGA	yBE 10i1 Bottom
SM10 Region 2	CACTCAATTATTTTATGTATATGTATGCT	yBE 10i2 Top
	AGCATACATATACATAAATATAATTGAGTG	yBE 10i2 Bottom
SM10 Region 3	ATGCTCAAAATCACCTGCCAATAACCCTGCAGG	yBE 10i3 Top
	CCTGCAGGGTTATTGGCAGGTGATTTTGAGCAT	yBE 10i3 Bottom
SM10 Region 1 Scan Mutant	gCtTaCaTgTgGcAcTgTgAgGgAcCcCgC	y10i1 T Scrm
	GcGgGgTcCcTcAcAgTgCcAcAtGtAaGc	y10i1 B Scrm
SM10 Region 3 Scan Mutant	cTtCgCcAcAgCcCaTtCaAcTcAaCaTtCcGt	y10i3 T Scrm
	aCgGaAtGtTgAgTtGaAtGgGcTgTgGcGaAg	y10i3 B Scrm
SM10 Region 1 sub1	gggCgggCgggCAAATTTTATGTAACACTC	y10i1 sub1 T
	GAGTGTTACATAAAATTTGcccGcccGccc	y10i1 sub1 B
SM10 Region 1 sub2	TCGTCCgggCgggCgggCTATGTAACACTC	y10i1 sub2 T
	GAGTGTTACATAGcccGcccGcccGGACGA	y10i1 sub2 B
SM10 Region 1 sub3	TCGTCCCTTTTGgggCgggCgggCACACTC	y10i1 sub3 T
	GAGTGTGcccGcccGcccCAAAGGGACGA	y10i1 sub3 B
SM10 Region 1 sub4	TCGTCCCTTTTGAATTTTgggCgggCgggC	y10i1 sub4 T
	GcccGcccGcccAAATTTCAAAGGGACGA	y10i1 sub4 B

**Table S5. Oligonucleotides for cloning *bab1* and *bab2* shRNAs into *NheI* and *EcoRI* sites of pattB-NE3 vector.**

siRNA name and sequence	Oligo name	Oligo Sequence (5' – 3')
bab1 siRNA 3 TTCGTCTGATAGTTGTTCCAG	b1_3 Top	ctagcagtCTGGAACAACAATCAGACGTAtagttatattcaagcataTTCGTCTGATAGTTGTTCCAGcg
	b1_3 Bottom	aattcgcCTGGAACAACATATCAGACGAAatgcttgaatataactaTACGTCTGATTGTTGTTCCAGactg
bab1 siRNA 4 TACAGCATGACCTTGACTCTC	b1_4 Top	ctagcagtGAGAGTCAAGCTCATGCTGAAtagttatattcaagcataTACAGCATGACCTTGACTCTCcg
	b1_4 bottom	aattcgcGAGAGTCAAGGTCATGCTGTAtatgcttgaatataactaTTCAGCATGAGCTTGACTCTCactg
bab2 siRNA 16 TATTTCAAAGTCCACAATCTG	b2_16 Top	ctagcagtCAGATTGTGGTCTTTGAAAAAtagttatattcaagcataTATTTCAAAGTCCACAATCTGcg
	b2_16 Bottom	aattcgcCAGATTGTGGACTTTGAAATAtatgcttgaatataactaTTTTTCAAGACCACAATCTGactg
bab2 siRNA 12 TTGGACTTCACCAGCTCCGTT	b2_12 Top	ctagcagtAACGGAGCTGCTGAAGTCCTAtagttatattcaagcataTTGGACTTCACCAGCTCCGTTcg
	b2_12 Bottom	aattcgcAACGGAGCTGGTGAAGTCCAAatgcttgaatataactaTAGGACTTCAGCAGCTCCGTTactg
bab2 siRNA 20 TCGAACTGATCGATTTCGCCG	b2_20 Top	ctagcagtCGGCGAAATCCATCAGTTCCAtagttatattcaagcataTCGAACTGATCGATTTCGCCGcg
	b2_20 Bottom	aattcgcCGGCGAAATCGATCAGTTCGAtatgcttgaatataactaTGGAAGTATGATGATTTCGCCGactg