Title

Bioinformatic and experimental analysis of hox and epidermal wound response enhancers in Drosophila

Author

Pearson, Joseph Carlisle

Publication Date

2007

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Bioinformatic and Experimental Analysis of Hox and Epidermal Wound Response Enhancers in *Drosophila*

A Dissertation submitted in partial satisfaction of the requirements for the degree

Doctor of Philosophy

in

Biology

by

Joseph Carlisle Pearson

Committee in charge:

Professor William McGinnis, Chair
Professor John Huelsenbeck
Professor Pavel Pevzner
Professor James Posakony
Professor Steven Wasserman

2007

The Dissertation of Joseph Carlisle Pearson is approved,

and it is acceptable in quality and form for publication on microfilm:

____

____

____

____

____

Chair

University of California, San Diego

2007

For Cathy, my most passionate supporter,

And Charlie, my favorite distraction.

"Any scientist who couldn’t explain to an eight-year-old

what he was doing was a charlatan."

“Tiger got to hunt, Bird got to fly;

Man got to sit and wonder, “Why, why, why?”

Tiger got to sleep, Birg got to lang;

Man got to tell himself he understand.”

Kurt Vonnegut, Cat's Cradle

“So it goes.”

Kurt Vonnegut, Slaughterhouse-Five

TABLE OF CONTENTS

Signature Page..iii

Dedication...iv

Epigraph..v

Table of Contents...vi

List of Figures and Tables...viii

Acknowledgements..x

Vita...xii

Abstract..xiii

Chapter I. Discovering and Dissecting *Cis*-Regulatory Elements;

Hox Regulation of Developmental Genes..1

Introduction...2

In Vitro and *In Vivo* Identification and

Analysis of *Cis*-Regulatory Elements.......................................5

In Silico Methods for Cis-Regulatory

Element Discovery in *D. melanogaster*... 9

Hox regulation of developmental target genes....................................13

Figures...27

Chapter II. The Embryonic Limb Enhancer for *Dll* Requires Multiple

***In Silico*-Identified Motifs for Activation**.......................................34

Introduction..35

Results..41

Discussion..46

Materials and Methods...52

Figures..55

Chapter III. Common *Cis*-Regulatory Logic of the

***Drosophila* Wound Response**..58

Introduction..59

Results..63

Discussion..74

Materials and Methods...83

Figures..87

Chapter IV: Conclusion/Final Thoughts..100

Appendix A: TWINE: A Java Program for Simple

Graphically-Assisted Searches of Repeated and

Conserved *Cis*-Regulatory Motifs...106

Appendix B: Alignments and Annotations of *Cis*-Regulatory

Elements and *Drosophila* Orthologs..116

References...158

LIST OF FIGURES AND TABLES

Chapter I

Figure 1. Hox expression, genomic organization, and Hox binding sequences..27

Table 1. Direct Hox-regulated genes: Caenorhabditis elegans and Drosophila melanogaster.........................29

Table 2. Direct Hox-regulated genes: Xenopus laevis, mouse and human......30

Figure 2. Structures of representative Hox-response enhancers......................31

Chapter II

Figure 3. Multiple repeated motifs that are conserved in *Drosophila Dll304* are required for activation................55

Figure 4. *A. gambiae Dll* upstream sequences do not drive limb expression in *D. melanogaster*...............................57

Chapter III

Figure 5. Conserved *cis*-regulatory sequences upstream of *Ddc* require GRH consensus sites and ERK for wound-dependent activation..87

Figure 6. Minimal sequence requirements for Ddc WRE..............................88

Figure 7. Sequences other than AFB and GRH consensus sites contribute to Ddc WRE 90

Figure 8. AFB and GRH consensus sites are required for *ple* WRE.............92

Figure 9. *Ddc*, *ple*, *msn*, and *kkv* are rapidly transcribed after wounding.......94

Figure 10. Conserved AFB and GRH consensus

site clusters identify *kkv* and *msn* WREs..95

Figure 11. AFB and GRH consensus site

requirements for *kkv1* and *msnSubB* WREs..97

Figure 12. GRH binds required *Ddc* and *ple* GRH consensus sites...............98

Figure 13. WREs are differentially active in *grh*^{IM} mutants..........................99

Figure 14. TWINE main window...114

ACKNOWLEDGEMENT

The first person I should, and thus will, thank, is Dr. William McGinnis, who has had unbelievable patience waiting for me to develop as a scientist. He welcomed me into his lab even though I was inexperienced, oblivious, and lazy. He deftly "hinted" at habits I might change in order to be a successful scientist. And as I progressed towards actually thinking critically about science, he has continued to challenge me intellectually, preparing me for the real world of academia.

The McGinnis lab, as a unit, is a surprisingly functional group of deviants, I can't imagine I would have survived in a less supportive lab. A special mahalo goes to the Mace/Ronshaugen pair, who took special care of me during my formative years. I am also thankful for the intellectual and material gifts from the Posakony lab over the years. Pass the pigs, please.

I am grateful for the financial support from the Cell, Molecular and Genetics Training Grant and the ARCS foundation.

The members of my committee have been very supportive and constructive in guiding my research, sagely moving me away from likely dead-ends, and graciously accepting Christmas sausage.

My family has always been patient, supportive, and loving towards me. I am a decent person because of my parents. My sister is a decent person despite of me.

To all teachers who, throughout my educational career, saw that I was lazy, and not just mediocre, thank you for making me feel guilty about not doing homework.

Cathy, for making me finally start studying, for coming to San Diego with me, and for keeping me motivated and happy every day, thank you. And thank you for Charlie.

Portions of Chapter I previously appeared in Pearson, J.C., Lemons, D., and McGinnis, W.J. (2005) Modulating Hox Gene Functions During Animal Body Patterning. *Nature Reviews Genetics* 6, 893-904. I was the primary author of the paper.

Portions of Chapter III previously appeared in Mace, K.A., Pearson, J.C., and McGinnis, W.J. (2005). An epidermal barrier wound repair pathway in *Drosophila* is mediated by *grainy head. Science* 308, 381-385. I was responsible for the research included in this dissertation.

Portions of Chapter III appear in Pearson, J.C., Juarez, M.T., Lidsky, I., Drivenes, O., and McGinnis, W.J. Common *Cis*-Regulatory Logic of the *Drosophila* Wound Response. *In Preparation.* I am primary author and responsible for the research included in this dissertation.

VITA

2001	Bachelor of Science, University of Redlands
2007	Doctor of Philosophy, University of California, San Diego

PUBLICATIONS

Mace, K.A., Pearson, J.C., and McGinnis, W.J. (2005). An epidermal barrier wound repair pathway in *Drosophila* is mediated by *grainy head. Science* 308, 381-385.

Pearson, J.C., Lemons, D., and McGinnis, W.J. (2005) Modulating Hox Gene Functions During Animal Body Patterning. *Nature Reviews Genetics* 6, 893-904.

ABSTRACT OF THE DISSERTATION

Bioinformatic and Experimental Analysis of Hox and

Epidermal Wound Response Enhancers in *Drosophila*

by

Joseph Carlisle Pearson

Doctor of Philosophy in Biology

University of California, San Diego, 2007

Professor William McGinnis, Chair

Unlike the well-known correspondence between the mRNA sequence blueprint and the protein encoded from it, fairly little is understood about how *cis*-regulatory elements, the DNA sequences that control when and where mRNAs are expressed, are structured to exert their influence. Even genes with well-conserved expression patterns in distantly related organisms, such as *Dll* in developing limbs of protostomes and deuterostomes, are controlled by largely unknown mechanisms. The number of techniques available for understanding *cis*-regulation is expanding rapidly, but each one has severe limitations. In an attempt to improve the rate of *cis*-regulatory discovery, I have combined molecular biological and *in silico* techniques to study two *cis*-regulatory paradigms in *Drosophila*. To dissect the *cis*-regulatory mechanisms controlling *Dll* limb expression in insects, I used germline transformation and bioinformatics to identify several novel motifs required for embryonic limb expression of *Dll.* Based on the techniques developed studying *Dll,* I have extended previous research dissecting a *cis*-regulatory element controlling *Ddc* wound-induced expression. I have identified a battery of wound-response genes, including the particular *cis*-regulatory elements controlling wound-responsive expression. These elements reveal complex regulatory interactions that result in the induction of a diverse set of genes required for various aspects of *Drosophila* wound healing.

Chapter I

Discovering and Dissecting *Cis*-Regulatory Elements;

Hox Regulation of Developmental Genes

Introduction

The human genome, at current count, contains between 20,000 and 25,000 genes (I.H.G.S.C., 2004). Pre-genomic era estimates placed the number closer to the 35,000 to over 100,000 genes (Lander *et al.*, 2001), based on our position at the pinnacle of evolution. The revised gene count is disquieting, especially when compared to the genomes of "simpler" organisms, such as *Drosophila melanogaster* with 14,601 genes (http://flybase.org/static_pages/docs/release_notes.html) or *C. elegans* with 20101 genes (http://www.wormbase.org/wiki/index.php/WS174).

Of course, gene count is such a simplistic and meaningless measure of genomic complexity that it serves little use other than as a simple statistic to cite in introductions, just as the idea of "Junk DNA" is primarily mentioned by researchers who study "Junk DNA" when they have new evidence demonstrating that "Junk DNA" performs essential functions. Cells do not express all genes, full-blast, at all stages of cellular and organismal development. Several levels of regulation tightly orchestrate the set of genes that is present at a given time in different cells, with additional regulatory mechanisms affecting gene product function.

Post-translational regulation of proteins, including cleavage, multimerization, covalent addition of molecules, and cellular localization increase the regulatory potential that can contribute to morphological and functional complexity. Similarly, post-transcriptional regulation mechanisms such as microRNA-based translation inhibition or transcript degradation, transcript localization, and alternative splicing regulate the levels and sequences of the proteins translated from these transcripts.

Multiple modes of regulation at the DNA level also operate as major components of determining the complement of mRNAs and proteins active in a given cell. The chromatin state, the configuration of Protein-DNA complexes along the genome as well as the set of chemical modifications to both, can determine whether genomic regions will be transcribed at all.

Perhaps the most important mechanism of gene regulation is the gene-specific activation or repression of transcription (Wray *et al.*, 2003). This mechanism involves the binding of combinations of sequence-specific DNA binding proteins (transcription factors) to specific DNA sequences near the regulated gene, called *cis*-regulatory elements. Depending on the set of transcription factors bound to these DNA regions, transcription at associated genes is increased or decreased. *Cis*-regulation is most likely the primary mechanism for controlling developmental, cell cycle, and environmentally induced changes in gene and protein expression.

If one considers the almost infinite different combinations of genes that could potentially be expressed in a given cell, combined with the different modifications that can alter protein structure and function, it becomes obvious how morphological complexity is not invariably proportional to gene number. Genetic networks are activated throughout development in different cell lineages, generating the myriad of specialized tissues. Changes in gene regulation, rather than changes in genes themselves, have been implicated in major morphological changes, such as *Drosophila* larval trichomes (Sucena and Stern, 2000; Sucena et al., 2003) and body pigmentation patterns (Jeong *et al.*, 2006), and even the incredible diversity of domesticated dog size (Sutter *et al.*, 2007). *Cis*-regulatory changes, rather than protein changes, are quite possibly the major driving force behind macroevolution (Rodriguez-Trelles *et al.*, 2003; Wray *et al.*, 2003).

***In vitro* and *In vivo* Identification and Analysis of *Cis*-Regulatory Elements**

Several methods have been developed for dissecting transcriptional regulation of gene expression, depending on goals of the researchers and available information about the genetic network and studied species. Given a gene of interest with a known expression pattern, potential upstream regulators with overlapping or complementary expression can be tested using mutants in the regulator genes to test for alteration of target gene expression. The reverse can also be used to test potential targets of a transcription factor of interest, by testing sets of genes that are expressed in patterns suggesting positive or negative regulation by the transcription factor. For example, *buttonhead* was identified as a regulator of *Distal-less* because of common expression in the ventral thorax (Estella *et al.*, 2003).

However, altered target expression in mutants for a given regulator does not necessarily indicate direct regulation. Identified regulators may activate or repress expression of intermediate genes that directly regulate the target gene. In such cases, mutants for the tested upstream regulator would still have altered target gene expression, without directly interacting with the target gene's *cis*-regulatory sequences. Additionally, classical mutants for developmentally-active transcriptional regulators have been identified largely because of the extreme, often lethal phenotype caused by mutations in these genes. These regulators control the expression of a wide variety of genes in multiple stages during development, and mutations in these regulators tend to have highly pleiotropic effects on development. Thus, it can be difficult to differentiate alterations in target gene expression due to changes in direct regulator interactions with *cis*-regulatory elements from alterations due to a fundamentally altered cellular identity. As an example by *reductio ad absurdum,* one could not claim that the transcription factor Ultrabithorax regulates a target gene expressed in the adult head simply because homozygotes for an embryonic lethal *Ubx* allele do not develop to adulthood, and thus have no adult head expression of the target gene.

Ectopic expression of regulators in a temporally and developmentally controlled manner, such as using the GAL4-UAS expression system, can avoid some of these problems. By limiting the axial breadth and developmental scope, fewer genes will be affected, and more precise conclusions can be made depending on co-incidence of expression patterns of ectopically expressed regulators and putative target genes. For example, over-expressing Deformed (DFD) in alternating embryonic segments in *D. melanogaster* ectopically activates *reaper* (*rpr*) mRNA expression in corresponding segments (Lohmann *et al.*, 2002), providing strong evidence that DFD activates *rpr* cell-autonomously, whether directly or indirectly. Additional potential issues can sometimes arise because of the often non-physiological expression levels of putative regulators when using ectopic expression systems, potentially spuriously activating genes that are not regulated under normal circumstances. Additionally, the proteins are being expressed in non-native cells that may not express other required cofactors that are present in cells that normally express the target gene of interest, potentially leading to a false negative result due to insufficiency.

Instead of studying gene regulation from the perspective of regulators, methods have been developed to identify and dissect the specific sequences within *cis*-regulatory DNA regions to which transcription factors bind and regulate transcription. Fragments of genomic DNA surrounding the gene of interest can be cloned into a vector containing a minimal promoter and a "reporter gene", such as the gene encoding β-galactosidase or Green Fluorescent Protein (GFP). When the proper combination of transcription factors binds to the regulatory DNA, in cell culture or *in vivo*, the reporter gene is transcribed and can be detected by *in situ* hybridization, by testing for enzymatic activity of the reporter gene product, or even by visualizing reporter prodcuts *in vivo*, such as in the case of GFP. Large regions of DNA containing a functional element can be subsequently split and subfragments individually tested to identify the smallest sequence that recapitulates native gene expression, thus defining the minimal *cis*-regulatory element.

Cis-regulatory elements are composed of a set of binding sites for transcription factors, generally including sites for both activators and repressors (sometimes through the same site) (Capovilla and Botas, 1998). Compared to protein-encoding DNA, very little is understood about *cis*-regulatory element structure and evolution. *Cis*-regulatory elements are generally assumed to be simply clusters of unordered binding sites (Arnosti, 2003), but cases of strict spacing requirements between motifs have been observed (Matsuo and Yasuda, 1992; Nikolajczyk *et al.*, 1996). *Cis*-regulatory elements for a given gene tend to exist as separate modules that independently control the various domains of expression of that gene.

Identifying the set of binding sites that are required for *cis*-regulatory function can be very informative, as this knowledge can lead to identification of the set of regulators that bind to these sites, thus elucidating the mechanism for controlling this gene expression. These binding sites can be identified by detecting direct interactions of the site and its binding protein, either *in vitro* (*e.g.* DNase I footprinting (Brenowitz *et al.*, 1986)) or *in vivo* (*e.g.* Chromatin Immuno-Precipitation). Candidate regulators can also be chosen by virtue of similarity of the required binding site to known binding sites of transcription factors.

One intriguing set of well-studied developmental regulators is the Hox family, the members of which specify Anterior-Posterior identity in animals. An apparent paradox arises between the incredible diversity of target genes that individual Hox members regulate with high specificity, and the low apparent *in vivo* specificity that Hox proteins have for particular DNA binding sequences within *cis*-regulatory elements. It is quite apparent that multiple levels of regulation must be involved in determining when and where Hox proteins exert their effects on transcription of target genes during development.

In Silico Methods for *Cis*-Regulatory Element Discovery in *D. melanogaster*

In addition to *in vitro* and *in vivo* exploration of *cis*- and *trans*-regulatory interactions, computer-based, or *in silico*, analysis of DNA sequences is an increasingly powerful tool for determining which DNA sequences are likely involved in *cis*-regulation. The rapid expansion in the amount of available genomic sequence of carefully selected groups of model organisms and their close relatives make *in silico* analyses of *cis*-regulatory elements much more powerful.

Non-coding sequences evolve much more quickly than protein-coding DNA, but required transcription factor binding sites are conserved to a much greater degree than surrounding non-coding sequences. Aligning homologous sequences spanning the *cis*-regulatory element reveals these conserved motifs. This technique is called "phylogenetic footprinting", due to its similarity to DNase I footprinting in revealing transcription factor binding sites, since conserved sequences were "protected" during evolution (Tagle *et al.*, 1988).

Clusters of binding sites matching the consensus binding site are highly likely to be *in vivo* targets of that transcription factor (Berman *et al.*, 2002; Markstein *et al.*, 2002), especially when looking for clusters of binding sites that are statistically unlikely to appear at random. Some transcription factors bind in a highly specific manner to fairly long DNA sequences, so random occurrence of non-functional sequences within the genome that happen to match these binding sites are rare. However, if the transcription factor's consensus sequence is small or weak (degenerate), it is much more difficult to recognize *bona fide* binding sites from the background of sequences that resemble binding sites simply by chance. Hox binding sites are members of this latter class, as the consensus binding sequence based on *in vivo* binding sites is essentially ATTA, nor does the degree of conservation strongly correlate with relative importance of the site to Hox regulation *in vivo*.

While genome-wide searches for statistically unlikely clusters of binding sites have revealed multiple *cis*-regulatory elements in different developmental contexts, the problem of "background noise" from non-functional sequences resembling true binding sites persists. Additionally, despite the tendency for multiple instances of binding sites to be present in functional *cis*-regulatory elements, these repeated motifs are again substantially masked by background sequences. Phylogenetic footprinting of homologous regulatory sequences has been particularly fruitful at a small scale, as constraining studied sequences to conserved regions defined by phylogenetic footprinting dramatically limits the search space to a relatively small cluster of distinct "islands" of putative binding sites. Alignments of *cis*-regulatory regions are much less constrained and tend to require significant manual input, while automated genome-scale alignments fundamentally misalign many intergenic and intronic sequences. Unfortunately, this means that genome-wide searches based on conserved regions derived from automated alignments are likely to miss a significant number of true matches.

As the number of sequenced genomes continues to expand, more robust alignment algorithms will continue to be developed that will more accurately reflect homology between sequences (GuhaThakurta, 2006). With computing power so cheap, it is entirely plausible to simply do parallel searches for a set of binding sites in multiple related genomes, and subsequently compare discovered clusters of even moderately over-represented sites for similarities in location relative to obviously conserved “anchor sequences” between genomes that establish relative location. Thus, linear conservation of particular sites is no longer an absolute requirement.

De novo discovery of binding sites within identified *cis*-regulatory elements without prior knowledge of likely regulatory factors is even more complicated, since the search must not only deal with spurious sequences similar to a defined consensus sequence, but must blindly retrieve the most over-represented DNA “words” from the set of all “words” in the element of sizes between 5 and 10 bp long, for example.

Compounded upon this is the ability of transcription factors to bind a set of sequences of varying degeneracy from the (unknown) consensus sequence, so any *in vivo* required sites can vary substantially, as long as it is bound strongly enough by the transcription factor so that its effects will be exerted. Thus, any putative degenerate matches to a given motif within the *cis*-regulatory element must be considered above some arbitrary threshold. Allowing too much degeneracy reduces the statistical difference of the observed number of matches to the expected number in random sequences, while too much stringency can exclude *in vivo* matches from consideration, thereby eliminating the motif from consideration because it didn’t have a statistically over-represented number of matches. A classic example of Scylla and Charybdis.

Phylogenetic footprinting dramatically improves *de novo* searches of over-represented motifs, for two reasons. First, it reduces the length of sequence is likely to be most important for *cis*-regulatory sequences, although not all functional sites are necessarily conserved (Ludwig *et al.*, 1998). Several algorithms now incorporate phylogenetic conservation into scoring of over-represented motifs within *cis*-regulatory elements (Sinha et al., 2004; Siddharthan et al., 2005).

Additionally, the pattern of evolutionary conservation often reflects DNA binding preferences of associated transcription factors, given a sufficient number of alignable genomes from which a binding matrix can be derived (Mirny and Gelfand, 2002). Thus, a reasonable theoretical binding matrix or consensus sequence can be generated from the set of similar conserved binding sites, and additional more degenerate matches to this matrix, conserved or not, can be more intelligently incorporated as additional site instances. Again, improved algorithms and increased availability of whole genome sequence is allowing massive-scale searches for novel *cis*-regulatory elements to take place (GuhaThakurta, 2006).

As more researchers take an integrated approach towards studying transcriptional regulation, combining *in silico, in vitro,* and *in vivo* methods, many new *cis*-regulatory elements are being discovered that regulate all manner of developmental and event-based transcription. While certain loose “rules” are emerging from these studies, it is increasingly clear that no single code analogous to codons exists (Wittkopp, 2006). Instead, a large set of overlapping patterns is emerging, and different *cis*-regulatory elements will reflect one or more of these patterns to drive transcription.

Hox regulation of developmental target genes

How has the evolution of animal genomes led to the amazing diversity of body forms that we observe in the natural world? Some of the most informative clues to this fundamental problem have come from the study of mutations in homeobox (Hox) genes. These mutations have powerful and interpretable effects on morphology, the most conspicuous being the homeotic transformations in *Drosophila melanogaster* (Lewis, 1978; Kaufman et al., 1990). Additionally, Hox genes are present and expressed in similar patterns in nearly every bilateral animal that has been analyzed, so their roles in morphological diversification probably evolved before the appearance of the first bilateral animal. Indeed, the initial glimpses into the conservation of metazoan developmental control genes came during the study of *D. melanogaster* Hox gene clusters (McGinnis and Krumlauf, 1992), which were originally (and more informatively) called homeotic selector genes.

In a wide variety of animals, ranging from nematodes to mice, mutations in Hox genes result in morphological defects that are restricted to discrete segmental zones along the anterior–posterior (A–P) axis, and sometimes include homeotic transformations similar to those that are seen in *D. melanogaster* (Beeman *et al.*, 1989; Krumlauf, 1994). Therefore, one conserved function of different members of the Hox gene family is to select one A–P axial identity over another. Hox genes are also interesting because their control of axial morphology has an abstract quality, exerting its influence in various organs, tissues and cell types within different A–P regions. Although emphasizing the role of Hox genes in controlling A–P or oral-aboral axial identities is a simplification of Hox gene functions, which have diversified during their 600 million years of evolution in millions of animal lineages (Bienz, 1994; Arenas-Mena et al., 1998; Ishii, 1999; Zakany and Duboule, 1999; Arenas-Mena et al., 2000), it is likely to be their ancestral role in developmental patterning (Finnerty *et al.*, 2004).

The Hox genes map in chromosomal clusters, and the different paralogs in the cluster are usually arranged in a collinear manner relative to their distinct, often overlapping, expression domains (Fig. 1a,b). In animal embryos in which mid-head and posterior abdomen can be distinguished, 'head' Hox genes have their initial anterior boundaries of expression in epidermal, neural and mesodermal cells of the mid-head region, and 'tail' Hox genes have their initial anterior boundaries of expression in the corresponding cell types of the posterior abdomen (McGinnis and Krumlauf, 1992). After the initial boundaries are set, Hox gene expression patterns can be labile within the larger confines of their initial domains (Castelli-Gair and Akam, 1995; Salser and Kenyon, 1996).

The homeodomain transcription factors that are encoded by the Hox genes activate and repress batteries of downstream genes by directly binding to DNA sequences in Hox-response enhancers. *In vitro*, Hox proteins can bind with high affinity as both monomers and multimers to specific DNA binding sites (McGinnis and Krumlauf, 1992) (excluding Labial/Homeobox 1 (LAB/HOX1) class proteins, which bind almost exclusively as heterodimers with Pre-B-cell homeobox/CEH-20 (PBC) class proteins (Chang, 1995; Mann and Chan, 1996). In vivo, however, Hox proteins bind to and regulate transcription through a broad collection of binding sites

(Fig. 1c). On many target enhancers, Hox proteins cooperatively bind to canonical heterodimer-binding sites (Chang, 1995; Mann and Chan, 1996) with members of the PBC family of homeodomain proteins (called PBX in mammals, EXD in D. melanogaster, CEH-20 and CEH-40 in Caenorhabditis elegans) (Van Auken, 2002) and binding sites for the HTH/MEIS super-family of homeodomain proteins (which include MEIS or PREP in mammals, HTH in *D. melanogaster* and UNC-62 in *C. elegans* (Van Auken, 2002) are frequently also found nearby. The functional regulatory complex that acts on some Hox-response elements therefore often involves HOX–PBC–MEIS heterotrimers (Mann and Affolter, 1998). In part through the different binding preferences of distinct Hox proteins in these heterotrimer complexes, and in part through PBC/MEIS-independent mechanisms, distinct but overlapping combinations of downstream genes are activated and repressed, with the result being morphological diversity in axial domains.

Hox targets and morphological diversification

In part, Hox proteins act as high-level executives, regulating other executive genes (including themselves, *extradenticle* (*exd*) and *homothorax* (*hth*) (Kuziora and McGinnis, 1988; Popperl, 1995; Gould et al., 1997; Azpiazu and Morata, 1998; Henderson and Andrew, 2000)) that encode transcription factors or morphogen signals. However, there is accumulating evidence that they act directly at many other levels (Weatherbee *et al.*, 1998), even on the 'blue collar' genes that mediate adhesion, cell division rates, cell death and cell movement. It is often lamented in print that few

Hox target genes are known, but this is not true. There are at least 35 target genes, in a variety of organisms, for which there is good evidence for direct regulation by one or more Hox proteins (Tables 1,2). In addition to these well-characterized direct targets, many other genes are influenced by Hox expression but they have not been shown to be regulated directly by Hox genes. Recent microarray experiments have identified an even larger pool of potential target genes (Cobb and Duboule, 2005; Lei *et al.*, 2005; Williams, 2005).

Hox regulation: executive level.

There are many examples of direct Hox regulation of genes that encode cell–cell signalling molecules or other transcription factors. Many of these target genes were suggested as potential targets because their mutant phenotypes showed similarities to Hox mutant phenotypes. Others were suggested because their A–P expression patterns either mimicked or complemented the patterns of one or more Hox proteins, consistent with positive or negative regulation, respectively.

One executive target gene is *decapentaplegic* (*dpp*), which is expressed in an A–P domain of visceral mesoderm in *D. melanogaster*. This *dpp* expression pattern is provided, in part, by the Hox proteins Ultrabithorax (UBX) and Abdominal-A (ABD-A, which activate and repress *dpp* transcription, respectively (Capovilla and Botas, 1998). The localized production of DPP, a secreted morphogen of the bone morphogenetic protein (BMP) class, then triggers cell shape changes in the gut that are required for normal visceral morphology (Bienz, 1994). UBX and ABD-A also directly repress the *Distal-less* (*Dll*) gene in the *D. melanogaster* abdominal epidermis (Vachon, 1992) (note that Hox proteins can operate either as transcriptional activators, as UBX does on *dpp* in the visceral mesoderm, or as repressors, as UBX does on *Dll* in the epidermis). The *Dll* gene encodes a homeodomain transcription factor that promotes appendage development, so its repression by UBX results in an absence of limbs from the abdomen. In *C. elegans*, the gene that encodes the Twist transcription factor homologue, *helix-loop-helix 8* (*hlh-8*), is directly activated in mid-body mesodermal cells by the Hox proteins abnormal cell lineage 39 (LIN-39) and male abnormal 5 (MAB-5) (Liu and Fire, 2000) (Fig. 1a,b). The *hlh-8* gene is required for normal mesoderm development, and its absence contributes to the localized muscle defects that are observed in *lin-39* and *mab-5* mutants.

Hox regulation: cell adhesion.

It has been long realized that Hox proteins must regulate cell adhesion, division, death, migration and shape in order to mould morphology (Garcia-Bellido, 1977). However, we have only recently learned the identities of some of the Hox target genes, the realizator genes (Garcia-Bellido, 1977), that directly mediate such properties at the cellular level in developing animals. Some of the first evidence for Hox control of cell adhesion came from Yokouchi (Yokouchi, 1995). Mouse *Hoxa13* is normally expressed in developing autopods. Ectopic activation of *Hoxa13* throughout the entire developing limb resulted in a marked reduction of the cartilage primordia for the proximal limb, cartilage that would normally develop into the radius and ulna (Yokouchi, 1995). This phenotype was associated with a *Hoxa13*-dependent increase in homophilic cell adhesion in proximal cartilage primordia.

Conversely, in mouse *Hoxa13* mutants, the mesenchymal condensations that normally form in the autopod are loosely and poorly organized, resulting in loss or abnormalities of the digit, carpal and tarsal bones that derive from the distal limb (Stadler *et al.*, 2001). Normally the gene that encodes the ephrin receptor EPHA7 is expressed in distal limb domains in a way that closely matches *Hoxa13* expression. However, in *Hoxa13* mutants *EphA7* expression is severely reduced. Reducing EPHA7 protein function with blocking antibodies in a *Hoxa13* background results in a failure to form the normal chondrogenic condensations in distal limb primordia, similar to the phenotype that is seen in *Hoxa13* mutants. Since in many contexts, direct interactions between transmembrane ephrin receptors and their membrane-bound ligands are required for normal cell adhesion (as well as for many other cellular responses) (Poliakov *et al.*, 2004), it seems likely that *Hoxa13*-mediated mesenchymal condensations in the distal limb are achieved in part by the activation of *EphA7* gene expression.

The regulation of ephrin receptor and/or ephrin ligand genes by Hox proteins seems to be common. In combination with PBX1, the HOXA1 and HOXB1 proteins can bind to and activate a mouse rhombomere-specific enhancer from the *Epha2* gene in COS7 CELLS (Chen and Ruley, 1998), and mouse HOXA9 protein can bind and activate the *Ephb4* gene in cultured endothelial cells (Bruhl, 2004). In addition, a recent genomic screen for Hox target genes has revealed that the mouse *Epha3* gene is repressed in a *Hoxd13*-dependent and *Hoxa13*-dependent manner in the posterior regions of developing autopods (Bromleigh and Freedman, 2000).

Hox regulation: cell cycle.

There is ample evidence for Hox involvement in blood cell development in mammals (Magli *et al.*, 1997), including the activation of *Hoxa10* gene expression during the differentiation of cultured myelomonocytic cells into monocytes. The role of *Hoxa10* in myeloid and erythroid development in bone marrow cells is complex, and it is not clear how well its function in cultured myelomonocytes recapitulates its function in animals (Thorsteinsdottir, 1997). With that caveat, forced expression of HOXA10 protein in cultured myelomonocytic cells results in premature differentiation into monocytes, accompanied by growth arrest (Bromleigh and Freedman, 2000). This growth-arrest phenotype seems to be controlled by *Hoxa10*-dependent activation of the *Cdkn1a* gene, which encodes a cyclin dependent kinase inhibitor, p21. The HOXA10 protein, together with the PBX1 and MEIS1 proteins, can bind *Cdkn1a* promoter sequences in vitro, which are presumably part of the cis-regulatory DNA that mediates the effects of HOXA10 on the cell cycle in vivo.

Hox regulation: cell death.

Another way in which Hox proteins might regulate morphology would be simply to ablate cells that are not part of the desired tissue shape. There is indeed evidence for Hox genes acting as sculptors by regulating cell death. In *D.*

melanogaster embryos, maintenance of the segmental boundary between the maxillary and mandibular segments of the head (Fig. 1a) requires localized cell death at the boundary that is controlled by the apoptosis-promoting gene *reaper* (*rpr*). Mutants in the Hox gene *Deformed* (*Dfd*) have a similar head segmental defect to mutants with a deletion for several cell death genes, and this is mainly due to the absence of *rpr* expression at the maxillary–mandibular border in *Dfd* mutants (Lohmann *et al.*, 2002). When a stripe of *rpr* expression is provided at the border in *Dfd* mutants, the segmental boundary is maintained. Additionally, a small *rpr* enhancer was defined that requires four DFD-binding sites for transcriptional activation at the maxillary–mandibular border in embryos (Lohmann *et al.*, 2002) (Fig. 2d).

Similarly, the morphology of the abdominal region of the *D. melanogaster* adult CNS is sculpted in a Hox dependent manner. In adults, the abdominal CNS is much smaller than the thoracic CNS, owing to fewer cells. Bello *et al.* found that a brief pulse of ABD-A protein expression in a large subset of the abdominal postembryonic neuroblasts triggers apoptosis in a manner that is dependent on the pro-apoptotic genes *rpr*, *head involution defective 1* (*hid1*) and grim*,* with a consequent size reduction of the adult abdominal neuromeres (Bello *et al.*, 2003).

<u>Hox regulation: cell migration.</u>

Hox genes have long been known to modulate cell migration, and one of the best examples of this activity is the control of Q neuroblast migration during *C. elegans* development by the Hox genes *mab-5* and *lin-39*. The function of *mab-5* is required cell autonomously for the posterior migration of the descendants of the QL neuroblast (Salser and Kenyon, 1992), and *lin-39* is required for the anterior migration of the descendants of the QR neuroblast (Clark *et al.*, 1993; Wang, 1993). The cell biological mediators that are regulated by the LIN-39 and MAB-5 proteins are not known.

The above examples barely scratch the surface of Hox-regulated morphological effector genes, yet they indicate that the cell biological effectors that are regulated by Hox proteins to sculpt morphology on the A–P axis are highly diverse. Hox proteins activate and repress multiple effector genes, in diverse cell types and tissues, throughout embryonic development. Because of the immense complexity of these interactions, it is unlikely that we will ever completely understand, at the molecular level, how Hox genes define an entire segment to have thoracic as opposed to abdominal identity. We will have to settle with understanding how cellular adhesion and other properties are controlled by the Hox system at a smaller scale in the diversification of axial morphologies.

Hox-regulated enhancers

Although we might never have a complete picture of the Hox-dependent cell-biological changes that differentiate one segment from another, it is plausible that one day we will understand the principles on which Hox target enhancers are built, at least well enough to predict their locations in the genome at a reasonable frequency. Our definition of Hox target enhancers in this review includes only those with strong evidence for direct regulation by a Hox protein in developing animals. The most rigorous test for validating a direct target element, the 'gold standard', is to subtly mutate the Hox-binding sites of an enhancer so that they prefer to bind Bicoid, a non-Hox homeodomain protein. This change results in the enhancer having reduced binding affinity and therefore a reduced response to the putative Hox trans-regulator. Compensatory mutations are then introduced into the DNA-binding domain of the putative Hox trans-regulator that allow it to bind with high affinity to the mutant sites in the enhancer. If the altered protein regains the ability to regulate the altered enhancer, it is strong evidence that a specific Hox protein is binding to a specific enhancer in embryonic cells. Only a few Hox-regulated enhancers have been validated using this rigorous test (Sun et al., 1995; Haerry and Gehring, 1997; Capovilla and Botas, 1998) (Schier and Gehring, 1992; Capovilla et al., 2001), which has so far only been attempted in *Drosophila* embryos. However, as is typical for most *in vivo* enhancer studies in animals, it has been more common to test whether a mutant enhancer in which all Hox-binding sites were eliminated mimics the activity of the wild-type enhancer in mutant embryos that lack the predicted Hox trans-regulator (Fig. 2).

Common principles of Hox target enhancers.

The five enhancers that are shown in Figure 2 represent a sample of diverse Hox-responsive DNA elements. Although they differ in many ways, including organism of origin, they also share several properties.

One common property is tissue specificity. For example, the UBX-dependent enhancer from Drosophila *dpp* is active only in the visceral mesoderm (Fig. 2), and is inactive in the epidermal, CNS and somatic mesoderm cells that also contain UBX protein. This specificity is due to the *dpp* enhancer also being regulated by Biniou/FOXF, a visceral-mesoderm-specific forkhead-type transcription factor (Zaffran *et al.*, 2001). Two autoactivation enhancers from the *Dfd* gene also exemplify this 'tissue-specificity rule'. One, which maps 5 kb upstream of the *Dfd* transcriptional start site, is active only in the epidermal cells that express DFD protein at the maxillary–mandibular border (Zeng *et al.*, 1994). Although DFD protein also autoactivates *Dfd* transcription in the CNS, this process is mediated through another enhancer that maps to the large intron of the *Dfd* gene (Lou *et al.*, 1995).

A second common property of Hox-response elements is the requirement for multiple Hox-monomer-binding sites (Fig. 1c, Fig. 2), most of which possess an ATTA (or TAAT) core sequence. Many Hox-response elements also require Hox–PBC-heterodimer-binding sites (Fig. 2), and often contain MEIS-binding sites as well, at variably spaced distances from the Hox–PBC sites. The range of both Hox-monomer-binding and Hox–PBC-binding sequences is broad (Fig. 1c). This is consistent with the evidence indicating that there is no systematic relationship between the affinities for monomer or heterodimer sites in vitro and their functional importance in vivo (Appel and Sakonju, 1993; Grieder et al., 1997; Capovilla and Botas, 1998; Galant et al., 2002; Ebner et al., 2005). How the functional specificity of Hox-regulated enhancers is strengthened without the help of PBC or MEIS sites is unknown, but it is not surprising that natural selective pressures will 'use' any available mechanism to generate meaningful Hox-enhancer expression patterns. On the basis of genetic evidence in *D. melanogaster*, there are at least two other evolutionarily conserved transcription factors, Teashirt and Disco, that probably operate as Hox cofactors in specifying A–P axial identity (Fasano, 1991; de Zulueta et al., 1994; Mahaffey et al., 2001; Robertson et al., 2004). Whether these two proteins mechanistically interact with Hox proteins to activate or repress target enhancers, and how they do so, is unknown.

In silico searches for Hox targets.

The best hope for identifying at least a subset of Hox-response elements by bioinformatic means is to search for genomic regions that are enriched for Hox, PBC and MEIS consensus sites. To test the utility of this strategy, Ebner *et al.* searched the *D. melanogaster* genome for canonical LAB–EXD-heterodimer-binding sequences within 40 base pairs of an HTH-consensus-binding sequence, and identified 30 genomic regions that met these requirements (Ebner *et al.*, 2005). The expression patterns of genes near to 16 of these loci were tested for overlap with the LAB expression pattern. Besides the *lab* autoregulatory enhancer (the source of the sequence motifs), only one other potential LAB-response element was identified. It mapped to the first intron of the *CG11339* gene, which encodes an actin-binding protein that is activated in a LAB-like expression pattern in the endoderm (Ebner *et al.*, 2005).

Tests of a 2 kb genomic fragment that contains the LAB–EXD–HTH consensus indicated that it did not function as a Hox-response element. However, the authors tested other DNA fragments around the *CG11339* transcription unit and identified an upstream fragment that acted as a LAB-dependent enhancer when fused to a reporter gene. When tested with in vitro binding assays, this enhancer was found to possess an HTH-binding site as well as a LAB–EXD site. Interestingly, the latter was highly divergent from the canonical site that was used in the bioinformatic search, but is still required for enhancer activation in vivo. This LAB–EXD site also bound LAB protein as a monomer, contesting the prevailing belief that LAB had little or no DNA-binding affinity in the absence of EXD (Chan *et al.*, 1996a). It is possible that *CG11339* was identified as a LAB-responsive gene by accident, albeit an accidental find that led to interesting new insight concerning in vivo LAB–EXD regulation (Ebner *et al.*, 2005). In any case, the results of this study do not bode well for bioinformatic predictions of naturally evolved Hox–PBC response elements that use the current version of the 'DNA-binding-selectivity model' (Chan *et al.*, 1996b; Ryoo and Mann, 1999).

On the basis of the current body of knowledge, it is clear that Hox target elements do not observe simple rules. Even individual enhancers seem to be regulated by both PBC-dependent and PBC-independent mechanisms(Gould *et al.*, 1997). Given the great diversity in Hox-response enhancer structures, it seems that even modest success in predicting Hox-response elements will require more knowledge about the range of Hox protein interactions with cofactors and target DNA sites.

We have discussed recent advances in four areas of Hox regulatory biology. Although a few Hox realizator genes have been identified that illustrate how Hox genes accomplish their function of sculpting variations on a basic segmental shape, many remain to be identified. We think it entirely plausible that the number of known Hox morphological effector genes will expand until almost every gene that can mediate cell adhesion, division, migration and so forth will be found to be directly regulated by Hox proteins in some developmental context.

Recent evidence has revealed a surprising lability in Hox protein functions during evolution, and this lability makes them the best current system for understanding how transcription-factor functions evolve in animals. As we have reviewed here, this lability might be facilitated by their ability to interact with a great range of binding sites within enhancers, either with or without cofactors from the PBC and MEIS families of proteins. From this perspective, the difficulty with coming to a general understanding of how different Hox proteins achieve their functional specificity might simply be due to their basic principles of operation. As Hox proteins operate in so many different cell types and developmental stages, selective pressure might have acted on their functions so that they will observe as few 'rules' as possible, allowing them to fit into nearly all developmental genetic circuits to tweak morphology. To look at this in anthropomorphic terms, it is amazing what the Hox proteins can accomplish when they let the tissue-specific transcription factors get the credit for making muscle, bone, skin and nerve.

Acknowledgment

Portions of Chapter I previously appeared in Pearson, J.C., Lemons, D., and McGinnis, W.J. (2005) Modulating Hox Gene Functions During Animal Body Patterning. *Nature Reviews Genetics* 6, 893-904. I was the primary author of the paper.

Figure 1. Hox expression, genomic organization, and Hox binding sequences. **(a)** The panel on the left shows a stage 13 *Drosophila melanogaster* embryo that has been colored in the schematic to indicate the approximate domains of transcription expression for all Hox genes except *proboscipedia* (*pb*) (Kosman, 2004). The segments are labelled (Md, mandibular; Mx, maxillary; Lb, labial; T1–T3, thoracic segments; A1–A9, abdominal segments). The panel on the right shows a mouse (*Mus musculus*) embryo, at embryonic day 12.5, with approximate Hox expression domains depicted on the head–tail axis of the embryo. The positions of hindbrain rhombomeres R1, R4 and R7 are labeled. In both diagrams the colors that denote the expression patterns of the Hox transcripts are color-coded to the genes in the Hox cluster diagrams shown in b. Anterior is to the left, dorsal is at the top. **(b)** A schematic of the Hox gene clusters (not to scale) in the genomes of *Caenorhabditis elegans*, *D. melanogaster* and *M. musculus*. Genes are colored to differentiate between Hox family members, and genes that are orthologous between clusters and species are labeled in the same color. In some cases, orthologous relationships are not clear (for example, *lin-39* in *C. elegans*). Genes are shown in the order in which they are found on the chromosomes but, for clarity, some non-Hox genes that are located within the clusters of nematode and fly genomes have been excluded. The positions of three non-Hox homeodomain genes, *zen*, *bcd* and *ftz*, are shown in the fly Hox cluster (grey boxes). Gene abbreviations: *ceh-13*, *C. elegans homeobox 13*; *lin-39*, *abnormal cell lineage-39*; *mab-5*, *male abnormal 5*; *egl-5, egg-laying defective 5*; *php-3, posterior Hox gene paralogue 3*; *nob-1, knob-like posterior*; *lab, labial*; *pb, proboscipedia*; *zen, zerknullt*; *bcd, bicoid*; *Dfd, Deformed*; *Scr, Sex combs reduced*; *ftz, fushi tarazu*; *Antp, Antennapedia*; *Ubx, Ultrabithorax*; *abd-A, abdominal-A*; *Abd-B, Abdominal-B*. **(c)** A compilation of *in vivo* DNA binding sequences arranged by the structural type of homeodomain that is encoded by the Hox genes. The three classes are Labial, Central, and Abdominal-B. The listed DNA binding sequences that are bound by Hox monomers and Pre-B-cell homeobox/CEH-20 (PBC)–Hox heterodimers are those that are required for the function of one or more Hox-response elements in developing mouse (Popperl, 1995; Maconochie, 1997; Safaei, 1997; Houghton and Rosenthal, 1999; Bromleigh and Freedman, 2000; Shi et al., 2001; Lampe et al., ; Serpente, 2005), fly (Vachon et al., 1992; Appel and Sakonju, 1993; Capovilla et al., 1994; Graba, 1995; Heuer et al., 1995; Sun et al., 1995; Chan, 1997; Grieder et al., 1997; Haerry and Gehring, 1997; Kremser, 1999; Bromleigh and Freedman, 2000; Capovilla et al., 2001; Zhou et al., 2001; Galant et al., 2002; Ebner et al., 2005; Hersh and Carroll, 2005) or nematode (Liu and Fire, 2000; Cui and Han, 2003). As no known HOX1-monomer-binding (mouse) or LAB-monomer-binding (fly) sites have been found to be functional *in vivo*, only PBC–LAB-heterodimer-binding sites are shown. Consensus logos were generated using all verified Hox-binding sites with WEBLOGO (Crooks *et al.*, 2004).

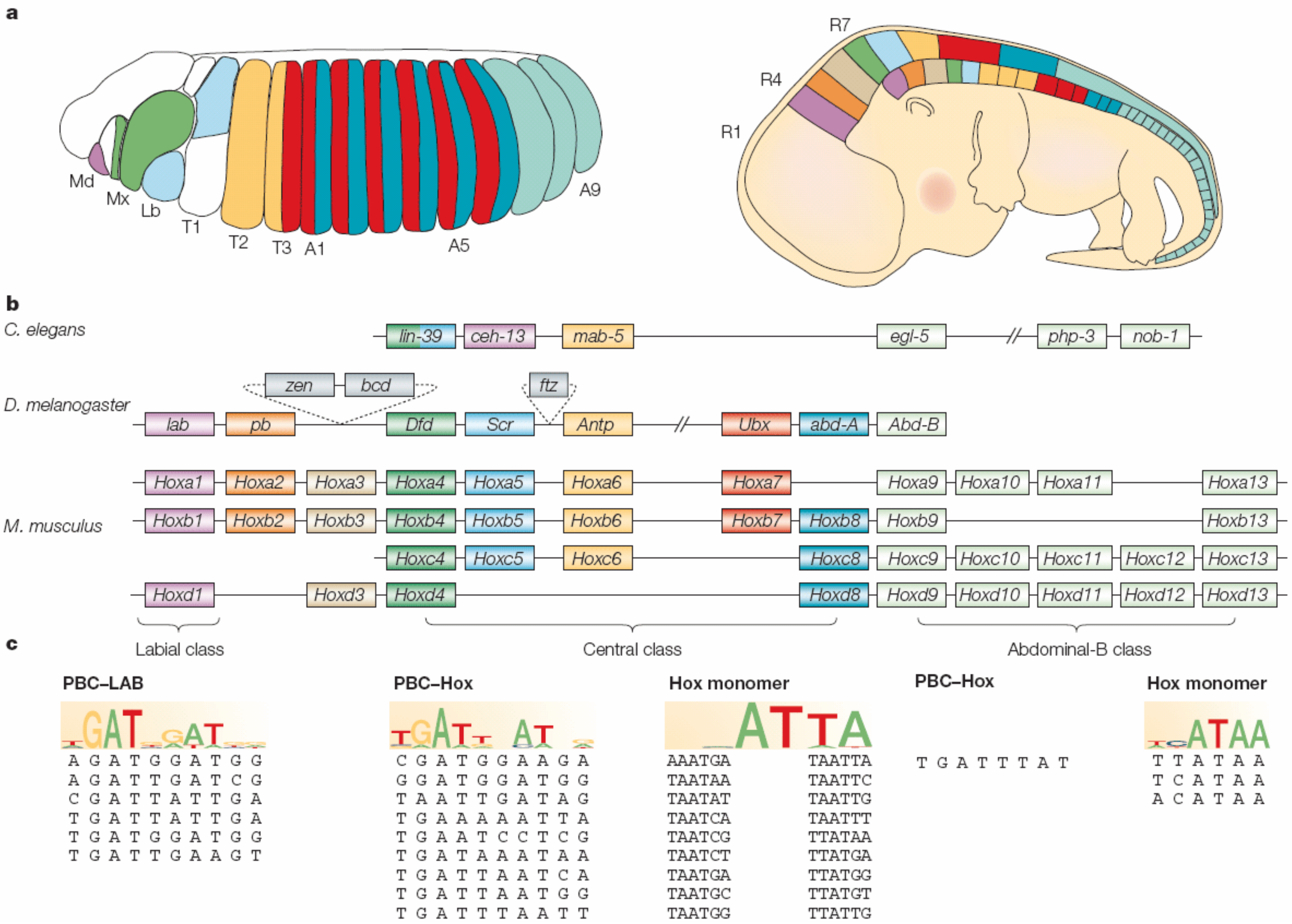

a
Md
Mx
Lb
T1
T2
T3
A1
A5
A9
R7
R4
R1
b
C. elegans
lin-39
ceh-13
mab-5
egl-5
php-3
nob-1
D. melanogaster
zen
bcd
ftz
lab
pb
Dfd
Scr
Antp
Ubx
abd-A
Abd-B
M. musculus
Hoxa1
Hoxa2
Hoxa3
Hoxa4
Hoxa5
Hoxa6
Hoxa7
Hoxa9
Hoxa10
Hoxa11
Hoxa13
Hoxb1
Hoxb2
Hoxb3
Hoxb4
Hoxb5
Hoxb6
Hoxb7
Hoxb8
Hoxb9
Hoxb13
Hoxc4
Hoxc5
Hoxc6
Hoxc8
Hoxc9
Hoxc10
Hoxc11
Hoxc12
Hoxc13
Hoxd1
Hoxd3
Hoxd4
Hoxd8
Hoxd9
Hoxd10
Hoxd11
Hoxd12
Hoxd13
Labial class
Central class
Abdominal-B class
c
PBC–LAB
A G A T G G A T G G
A G A T T G A T C G
C G A T T A T T G A
T G A T T A T T G A
T G A T G G A T G G
T G A T T G A A G T
PBC–Hox
C G A T G G A A G A
G G A T G G A T G G
T A A T T G A T A G
T G A A A A A T T A
T G A A T C C T C G
T G A T A A A T A A
T G A T T A A T C A
T G A T T A A T G G
T G A T T T A A T T
Hox monomer
AAATGA
TAATAA
TAATAT
TAATCA
TAATCG
TAATCT
TAATGA
TAATGC
TAATGG
TAATTA
TAATTC
TAATTG
TAATTT
TTATAA
TTATGA
TTATGG
TTATGT
TTATTG
PBC–Hox
T G A T T T A T
Hox monomer
T T A T A A
T C A T A A
A C A T A A

Table 1. Direct Hox-regulated genes: *Drosophila melanogaster*

Sorted by tissue type. ABD-A, Abdominal A; ANTP, Antennapedia; ChIP, chromatin immunoprecipitation; DFD, Deformed; LAB, Labial; SCR, Sex combs reduced; UBX, Ultrabithorax.

Regulated Gene	**Expression domain controlled by Hox**	**Regulating Hox protein(s)**	**Strongest evidence for direct Hox regulation**	**References**
forkhead	Embryonic salivary gland	SCR	Enhancer with mutated Hox site	(Ryoo and Mann, 1999; Zhou *et al.*, 2001)
Distal-less	Embryonic ectoderm	UBX, ABD-A	Enhancers with mutated Hox sites	(Vachon *et al.*, 1992)
Antp	Embryonic tracheal and neural ectoderm	ANTP, UBX, ABD-A	Enhancers with mutated Hox sites	(Appel and Sakonju, 1993)
Hoxa4	Embryonic epidermis	UBX	Bicoid site swap (K50) using UBX	(Haerry and Gehring, 1997)
Deformed	Embryonic maxillary epidermis	DFD	Enhancer with mutated Hox site	(Zeng *et al.*, 1994)
1.28	Embryonic maxillary epidermis	DFD	Enhancer with mutated Hox sites	(Pederson, 2000)
teashirt	Embryonic epidermis and somatic mesoderm	ANTP, UBX	Enhancers with deleted Hox sites	(McCormick *et al.*, 1995)
scabrous	Embryonic ectoderm	UBX, ABD-A, ABD-B	ChIP using UBX	(Graba, 1992)
Transcript 48	Embryonic epidermis, and somatic and visceral mesoderm	ABD-A, UBX	ChIP using UBX	(Strutt and White, 1994)
La-related protein	Embryonic ectoderm, and somatic and visceral mesoderm	SCR, UBX	ChIP using UBX	(Chauvet, 2000)
centrosomin	Embryonic visceral mesoderm and CNS	ANTP, UBX, ABD-A	ChIP using ANTP	(Heuer *et al.*, 1995)
decapentaplegic	Embyonic midgut visceral mesoderm	ANTP, UBX, ABD-A	Bicoid site swap (K50) using UBX and ABD-A	(Capovilla et al., 1994; Manak et al., 1994; Sun et al., 1995; Capovilla and Botas, 1998)
apterous	Embryonic muscle mesoderm	ANTP	Bicoid site swap (K50) using ANTP	(Capovilla *et al.*, 2001)
connectin	Embryonic mesoderm	ABD-A, UBX	ChIP using UBX	(Gould and White, 1992)
serpent	Embryonic lateral mesoderm	UBX	One-hybrid assay using UBX	(Mastick *et al.*, 1995)
wingless	Embryonic visceral mesoderm	ABD-A	Enhancers with mutated or deleted Hox sites	(Grienenberger, 2003)
Wnt4	Embryonic visceral mesoderm	ANTP, UBX, ABD-A	ChIP using UBX	(Graba, 1995)
beta-tubulin at 60D	Embryonic visceral mesoderm	UBX	Enhancers with deleted Hox sites	(Kremser, 1999)
labial	Embryonic midgut endoderm	LAB	Enhancer with mutated Hox site	(Grieder *et al.*, 1997)
CG11339	Embryonic midgut endoderm	LAB	Enhancers with mutated Hox sites	(Ebner *et al.*, 2005)
spalt major	Wing imaginal discs	UBX	Enhancers with mutated Hox sites	(Galant *et al.*, 2002)
knot	Wing imaginal discs	UBX	Enhancers with mutated or deleted Hox sites	(Hersh and Carroll, 2005)

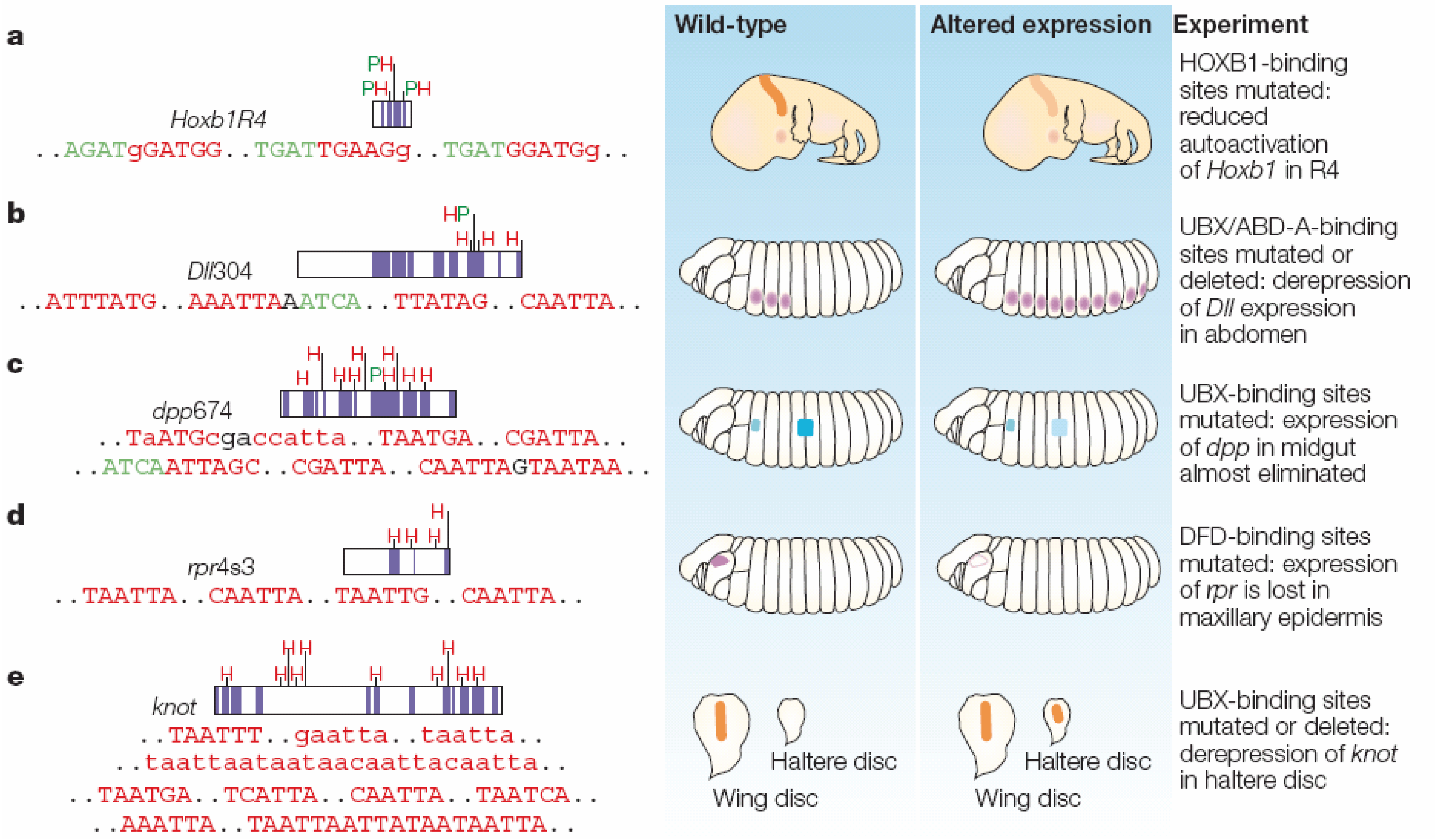
Table 2. Direct Hox-regulated genes: *Caenorhabditis elegans*, *Xenopus laevis*, mouse and human

Sorted by tissue type. *ceh-13*, *C. elegans homeobox gene 13*; ChIP, chromatin immunoprecipitation; *egl-17/18, egg-laying defective 17/18*; *elt-6, erythroid-like transcription factor family 6*; *hlh-8, helix-loop-helix 8*; LIN-39, abnormal cell lineage 39; MAB-5, male abnormal 5; R4, rhombomere 4.

Regulated gene	Expression domain controlled by Hox	Regulating Hox protein(s)	Strongest evidence for direct Hox regulation	References
C. elegans				
hlh-8	Larval M lineage cells	LIN-39, MAB-5	Enhancers with mutated Hox sites	(Liu and Fire, 2000)
egl-17	Primary vulval cells	LIN-39	Enhancer with mutated Hox site	(Cui and Han, 2003)
ceh-13	Embryonic dorsal body-wall muscle and ventral nerve cord	CEH-13	Enhancers with mutated Hox site	(Streit, 2002)
egl-18, elt-6	Larval vulval cells	LIN-39	Enhancers with mutated Hox sites	(Koh, 2002)
X. laevis				
Hoxb4, Hoxb5	Unspecified	HOXB4	Induced nuclear importation of HOXB4 after translation inhibition	(Hooiveld, 1999)
RAS-related protein-1a	Embryonic dorsal ectoderm?	HOXB4	Induced nuclear importation of HOXB4 after translation inhibition	(Morsi El-Kadi *et al.*, 2002)
iroquois 5	Embryonic neural ectoderm?	HOXB4	Induced nuclear importation of HOXB4 after translation inhibition	(Theokli *et al.*, 2003)
caspase-8-associated protein 2/FLASH	Embryonic notochord	HOXB4	Induced nuclear importation of HOXB4 after translation inhibition	(Morgan *et al.*, 2004)
M. musculus				
Hoxb1	R4	HOXB1	Enhancers with mutated Hox sites	(Popperl, 1995)
Hoxb2	R4	HOXA1, HOXB1	Enhancer with mutated Hox site	(Maconochie, 1997)
Hoxb3, Hoxb4	Hindbrain	HOXB4, HOXD4	Enhancers with mutated Hox sites	(Gould *et al.*, 1997)
retinoic acid receptor-beta	Embryonic hindbrain	HOXB4, HOXD4	Enhancers with mutated or deleted	(Serpente, 2005)
serine protease inhibitor 3	CNS	HOXB5	ChIP using HOXB5	(Safaei, 1997)
H. sapiens				
ephrin B4	Human umbilical venous endothelial-cell culture	HOXA9	ChIP using HOXA9, which was screened for Ephrin B4 by PCR	(Bruhl, 2004)

Figure 2. Structures of representative Hox-response enhancers. This figure illustrates five archetypal Hox-regulated enhancers from *Mus musculus* or *Drosophila melanogaster*. Enhancers are represented by white rectangles. Linear sequence conservation from *D. melanogaster* to *Drosophila virilis* (**b–e**) or from *M. musculus* to the puffer fish *Takifugu rubripes* (**a**) is represented by blue bars. Hox and Pre-B-cell homeobox/CEH-20 (PBC) sites that were identified by footprinting and/or mutation analysis are noted by H or P, respectively. Below the schematic of each enhancer are confirmed Hox or Hox–PBC binding sequences; Hox and PBC binding sites are colored in red or green text, respectively, and conserved sequences are capitalized. The wild-type expression pattern of each enhancer is shown on the right, where one example of evidence that confirms Hox dependence is described for each enhancer. **(a)** An enhancer that responds to both HOX and PBC proteins maps upstream of mouse *Hoxb1*. This enhancer contains a repeat of evolutionarily conserved HOXB1–PBX (Pre-B-cell homeobox)-heterodimer-binding sites that are required for autoactivation of *Hoxb1* in rhombomere 4 (R4)(Popperl, 1995). Other Hox-dependent enhancers with required canonical HOX–PBC-binding sites include those in *D. melanogaster labial* (Grieder *et al.*, 1997) and *forkhead* (Ryoo and Mann, 1999) and in *C. elegans helix-loop-helix* (*hlh8*)/*twist*(Liu and Fire, 2000). (**b**) An example of a Hox target that apparently requires Hox and Extradenticle (EXD) inputs through a non-canonical site is a thoracic-limb enhancer (*Dll304*) from the *Distal-less* (*Dll*) gene (Vachon *et al.*, 1992), which is repressed in the abdomen by Ultrabithorax (UBX) and Abdominal-A (ABD-A) through a repression element called DMX-R(Gebelein *et al.*, 2002; Gebelein *et al.*, 2004). DMX-R (panel b) has two Hox-binding sites (one in a non-canonical Hox–EXD-heterodimer site), as well as sites that bind a large multiprotein repression complex (Gebelein *et al.*, 2004). Curiously, when the non-canonical Hox–EXD site is changed to a canonical site with higher in vitro affinity, UBX and ABD-A no longer repress this *Dll* limb enhancer in vivo(Gebelein *et al.*, 2002). (**c**) An enhancer that is activated and repressed by different abdominal Hox proteins
The *dpp-674* enhancer of *decapentaplegic* controls expression in midgut primordia and is activated by UBX but repressed by ABD-A more posteriorly (Capovilla *et al.*, 1994). Eliminating the sites that ABD-A normally binds to repress transcription allows more posterior expression, whereas eliminating the sites that UBX binds almost eliminates expression in the midgut. Interestingly, a sub-element that lacks the repression Hox sites, but contains the activation sites, can be activated by either UBX or ABD-A(Capovilla and Botas, 1998). **(d)** Some Hox targets appear to be regulated independently of EXD. The ambiguity exists because it is often impossible to rigorously test Hox-response elements for dependence on PBC/*exd* due to the early developmental functions of zygotic or maternally contributed EXD protein (Peifer and Wieschaus, 1990; Rauskolb *et al.*, 1995). Evidence of *exd* independence/dependence is often limited to the absence/presence of EXD or EXD–HOX sites that can be identified by in vitro binding assays. By these criteria, two Hox-response elements that are activated by Deformed (DFD) in the maxillary epidermis are EXD-independent — a *1.28* enhancer and a *reaper* enhancer (Andrew *et al.*, 1994; Lohmann *et al.*, 2002; Pederson, 2000). (**e**) Other Hox-response elements are expressed in body regions

Figure 2. Structures of representative Hox-response enhancers (continued)
where EXD is not expressed. These include two wing-IMAGINAL-DISC enhancers that are directly repressed by UBX, one from the *knot* (Hersh and Carroll, 2005) gene and one from the *spalt major* gene (Galant *et al.*, 2002). Both are derepressed in the haltere imaginal disc in Ubx mutants, and both possess multiple UBX-binding sites that are required for the repression of the enhancers in haltere primordia.


a
Hoxb1R4
PH
PH
PH
..AGATgGATGG..TGATTGAAGg..TGATGGATGg..
b
Dll304
HP
H H H
..ATTTATG..AAATTAAATCA..TTATAG..CAATTA..
c
dpp674
H H H
H HH PH H H
..TaATGcgaccatta..TAATGA..CGATTA..
..ATCAATTAGC..CGATTA..CAATTAGTAATAA..
d
rpr4s3
H
H H H
..TAATTA..CAATTA..TAATTG..CAATTA..
e
knot
H H H
H HH H H H HH
..TAATTT..gaatta..taatta..
..taattaataataacaattacaatta..
..TAATGA..TCATTA..CAATTA..TAATCA..
..AAATTA..TAATTAATTATAATAATTA..
Wild-type
Altered expression
Experiment
HOXB1-binding sites mutated: reduced autoactivation of Hoxb1 in R4
UBX/ABD-A-binding sites mutated or deleted: derepression of Dll expression in abdomen
UBX-binding sites mutated: expression of dpp in midgut almost eliminated
DFD-binding sites mutated: expression of rpr is lost in maxillary epidermis
UBX-binding sites mutated or deleted: derepression of knot in haltere disc
Haltere disc
Wing disc
Haltere disc
Wing disc

Chapter II

The Embryonic Limb Enhancer for *Distal-less* Requires Multiple Novel *In Silico*-Identified Motifs for Activation

Introduction

Evolutionary change can be effected either by modifying the structure of proteins, thus affecting cellular function, or by modifying the expression of those proteins, altering where the proteins function. The former occurs when mutations occur in coding DNA, the latter in regulatory DNA. Several constraints are placed on whether mutations within coding DNA are tolerated. In order for a mutation in coding sequence to “survive” so that it can be established in a population, it generally cannot shift the reading frame, introduce premature stop codons, or adversely affect overall protein folding by amino acid changes or insertion/deletion events such that the function of the protein is inhibited. These rules help preserve the backbone of proteins as they evolve in different species, such that homologous proteins from species that diverged hundreds of millions of years ago can be identified by amino acid sequence. This makes it relatively easy to clone and then track and understand the evolution of homologous proteins in distantly related organisms.

The same evolutionary constraints are not found in regulatory DNA (Wittkopp, 2006; Wray *et al.*, 2003). Cis-regulatory elements seem to operate largely simply as clusters of binding sites that work as a binary code to determine whether a gene will be expressed in a given cell (“Billboard Model” (Arnosti, 2003)). Orientation and order of binding sites often doesn’t matter, nor does position or distance relative to the controlled gene (there are some constraints), and individual binding sites tend to work additively to modulate the strength of the effect of the binding protein on transcription. Thus, gain or loss of binding sites are not “make or break” situations that inevitably eliminate regulatory control of associated genes, but instead tweak the overall effect of the cis-regulatory element. For example, homologous cis-regulatory elements controlling *even-skipped (eve)* stripe 2 expression from different drosophilids all drive identical expression in *D. melanogaster* (except for slight quantitative differences) despite several mutations and even deletions of binding sites for known trans-acting factors (Ludwig *et al.*, 1998).

Additionally, transcription factors can bind to DNA sequences significantly diverged from the "consensus sequence" and still maintain *in vivo* function. The UBX-EXD heterodimer binding site found to confer the majority of abdominal repression of *Dll304*, the early embryonic limb *cis*-regulatory element, is ATTAAATCA. This differs from the canonical UBX-EXD binding site by the insertion of an additional nucleotide between the binding sites for UBX and EXD. Surprisingly, altering this binding site to the canonical ATTAATCA site in *Dll304* removes the repressive effect of UBX and EXD on the element (Gebelein *et al.*, 2002). This suggests that the set of sequences to which transcription factors bind *in vivo* differ depending on various contexts, and cannot be represented by simple consensus sequences or matrices, or by simply being limited the set of sites bound *in vitro*. Of course, it is also a formal possibility that transcription factors other than UBX and EXD are regulating *Dll* transcription through these sites.

The leeway allowed on *cis*-regulatory elements permits mutations (insertions/deletions, substitutions, shuffling of sites) to quickly accumulate without adversely affecting the ability of the element to properly control gene expression.

This means that homologous elements can very quickly become unrecognizable by sequence similarity, while continuing to be functionally identical. Beyond this point, homologous elements can only be identified by similar expression pattern (Bonneton *et al.*, 1997) or by finding statistically improbable clustering of shared binding sites (Berman *et al.*, 2002; Bonneton *et al.*, 1997; Markstein *et al.*, 2002; Rebeiz *et al.*, 2002). Unless the *cis*-regulatory element's inputs are already well-characterized, the latter method is difficult because of the background introduced by random DNA interspersed and around the functional element.

This means that, in order to understand cis-regulatory evolution beyond the small window of time where homologous elements are easily recognizable by sequence, the elements must be identified by their ability to drive similar expression patterns. Despite the inherent difficulty in identifying distantly related cis-regulatory elements, it is essential to dissect the evolution of expression of developmentally important genes, since modifying expression is probably a major driving force of evolutionary change.

Distal-less (*Dll*) encodes a homeodomain protein that primarily specifies body outgrowths, and is an ideal test case for attempting to unravel the mystery of how *cis*-regulatory elements evolve. *Dll* homologs are found in most invertebrates and vertebrates (Panganiban and Rubenstein, 2002), are expressed in similar areas of the body plan, and the requirement of *Dll* expression in the embryonic distal leg primordia has been confirmed even in spiders using RNAi (Schoppmeier and Damen, 2001).

Two discrete elements have been defined in *Drosophila melanogaster* that drive embryonic leg expression of *Dll*: *Dll304* initiates expression in thoracic spots beginning in early stage 11, and *Dll215* maintains expression in leg primordia and head structures through an auto-regulatory loop (Castelli-Gair and Akam, 1995; Vachon *et al.*, 1992). w*ingless* (*wg*) (Cohen *et al.*, 1993; Kubota *et al.*, 2003) and *buttonhead* (*btd*) (Estella *et al.*, 2003) are required for activation of *Dll304*, and Ultrabithorax (UBX) and Abdominal-A (ABD-A) both repress *Dll304* in the abdomen, by binding to at least two verified sites in the 3' part of *Dll304* (Vachon *et al.*, 1992; White *et al.*, 2000).

Dll function has apparently been maintained as an early developmental gene necessary for distal limb development since before the protostome-deuterostome split, since outgrowths such as limbs in both protostomes and deuterostomes express *Dll* during development. A simple hypothesis is that, at least in insect, a common set of regulators controls *Dll* expression in obviously homologous tissues such as legs. This is because it would be "simpler" to maintain the same mechanism of regulation as the ancestral metazoan, rather than develop novel methods of driving expression in distal leg primordia. If this is true, then homologous cis-regulatory elements that drive *Dll* expression in distal limb primordia in the native organism should provide qualitatively equivalent regulatory control if transferred into another insect, for example *D. melanogaster.*

A major issue in analysis of the mechanism of activating *Dll304* is identifying the controlling factors. The WG pathway was shown to have a positive input in initial transcriptional activation of *Dll* through the *Dll304* element (Cohen *et al.* 1993), and the Hox proteins UBX and ABD-A both repress the *Dll304* element through two sites, Bx1 and Bx2, located near the 3' end of *Dll304* (Vachon *et al.*, 1992). Antennapedia (ANTP), the Hox protein expressed in the thoracic segments, is not necessary for activation (Mann, 1994). Homothorax(HTH) and Extradenticle (EXD), as well as Engrailed (EN) and Sloppy paired (SLP), act as cofactors for the Hox proteins in *Dll304* repression (Gebelein *et al.*, 2004; White *et al.*, 2000), and both the DPP pathway and the EGF Receptor pathway is also involved in prevention of ectopic expression on the ventral-dorsal axis, although it is not clear if this is through repression of transcription or preventing cell migration (Goto and Hayashi, 1997). Ubx binds so indiscriminately to DNA that, although the official "consensus" sequence based on *in vitro* studies is CCATTAA, the functional consensus is essentially ATTA (Pearson *et al.*, 2005).

This low specificity by Ultrabithorax and lack of knowledge about whether known trans-acting factors are acting directly or by inducing transcription of direct activators makes binding site clustering search algorithms as implemented by Fly Enhancer (Markstein *et al.*, 2002), Cis-Analyst(Berman *et al.*, 2002), or SCORE (Rebeiz *et al.*, 2002) essentially useless. And even with a fairly well-defined element such as *Dll304* (877 bp), the background noise when performing a dot plot-type comparison against itself to discover repeated motifs is generally uninformative. This situation can be improved dramatically by eliminating the majority of the sequence by analyzing only conserved blocks. It is safe to assume that if a sequence of nucleotides has been conserved for 50 million years in multiple species, it serves some regulatory purpose. Using evolutionary conservation as a filter for functional sequences is known as phylogenetic footprinting (Tagle *et al.*, 1988), and footprinted sites in regulatory elements tend to have a good correlation with known binding sites for known controlling factors and can even be used to identify unknown inputs (Andrioli *et al.*, 2002; Kim, 2001; Ludwig *et al.*, 1998).

I used bioinformatic and molecular biological techniques to identify homologs to the *Dll304* limb regulatory element in other insects. I cloned homologs from several distantly related *Drosophila* species, and used phylogenetic footprinting to identify multiple repeated and conserved motifs within *Dll304.* I tested two of these novel motifs, confirming that they are indeed required for activation of *Dll304* in *D. melanogaster.* While multiple sequences have been identified that required for abdominal repression of *Dll304*, these motifs are the first required for activation.

Results

Dll304 is structurally and functionally conserved in *Drosophila*

Identification of important motifs within a *cis*-regulatory element can be greatly aided by phylogenetic footprinting, the comparison of homologous DNA sequences to identify conserved motifs. *D. virilis* is a commonly used species for *cis*-regulatory phylogenetic footprinting comparisons against *D. melanogaster.* Sufficient time has passed since divergence from *D. melanogaster* (~40 million years) (Russo *et al.*, 1995; Tamura *et al.*, 2004) for most neutral sequences to change in one or both species, while required coding and regulatory sequences are mostly maintained.

To identify the homolog to *D. melanogaster Dll304* (*DmDll304*), I screened a *D. virilis* genomic library using a probe of *DmDll304*, identifying several independent clones. I isolated and purified one of these clones, and identified by Southern Hybridization a single 1.8kb HindIII fragment to which the *DmDll304* probe bound. I subcloned and sequenced this fragment, its obvious homology revealing it as the homolog to *DmDll304* (figure 3a).

To test whether the identified *D. virilis* homolog *DvDll304* contained all sequences necessary for limb-specific expression, I cloned this fragment into the pH-Stinger GFP reporter vector and transformed *Drosophila* embryos (figure 3d). The resulting GFP expression recapitulated *DmDll304* expression (figure 3c), confirming that I had cloned the functional *D. virilis* homolog to *DmDll304*.

Because additional information can often be gained from phylogenetic footprinting by comparing multiple related species (phylogenetic shadowing) (Boffelli *et al.*, 2003), I cloned *Dll304* fragments from *D. hydei, D. immigrans,* and *Scaptodrosophila lebanonensis* by PCR using primers to conserved sequences between *D. melanogaster* and *D. virilis.* Additional flanking sequences were generated by inverse PCR. Additional homologs were identified from databases of whole genome shotgun sequencing of several drosophilids, and incorporated into an alignment of *Dll304* from widely diverged species of *Drosophila*.

Multiple i*n silico* identified motifs are required for *Dll304* activation

Alignment of all identified homologs of *Dll304* revealed several blocks of conservation containing putative binding sites for transcription factors (figure 3a). Analysis of the conservation profile of *Dll304* and flanking regions revealed that no sequences are linearly conserved in the first 300 base pairs (bp) of *DmDll304.* In contrast, several large blocks of conservation were identified beyond this point, even extending beyond the SspI site at 877bp that marks the end of the canonical *Dll304* sequence (Vachon *et al.*, 1992), to 971 bp. Since sequences from 300 bp to 731 bp are sufficient to confer expression similar to native *Dll* limb expression (Gebelein *et al.*, 2002; Gebelein *et al.*, 2004; Vachon *et al.*, 1992)*,* multiple independent *cis*-regulatory elements are likely to exist between 300 bp to 971 that regulate different aspects of *Dll* expression throughout development.

To identify important motifs contained within this conserved region, I compared all conserved sequences between *D. melanogaster* and *D. virilis* to itself using WinDotter dot plot comparison tool (Sonnhammer and Durbin, 1995). These comparisons revealed several motifs that are not only conserved in these distantly related drosophilids, but also repeated in multiple conserved positions in both species. I then checked all other cloned homologs for these identified motifs, generating a set of sequences that are repeated in all known *Drosophila Dll304* sequences. Searching for these motifs in all cloned species revealed that, in addition to the conserved motif instances that were initially discovered, several other matches were found that are only conserved in a subset of *Drosophila* species, similar to patterns seen in other robust *cis*-regulatory elements (Ludwig *et al.*, 1998).

Analysis of positions of conserved repeated motifs in *Dll304* revealed clustering of motifs to different parts of the studied DNA sequence. Several motifs are clustered in the 3' *Dm/Dv c*onserved block. AATTGACA is repeated thrice within 100 bp, all three instances almost perfectly conserved. A conserved palindrome, TTGCTTAAGCAA, is also found in this region. One conserved instance of motif MATAYTTGSGMAAWTAAAT is found in this region, with a second conserved instance in a more 5' conserved block within the minimal *Dll* limb element (*Dll304Min)*. Since the 3' region is not required for expression in embryonic limb spots, this set of motifs likely controls *Dll* expression in other tissues.

We identified two novel conserved motifs were repeated within the bounds of the minimal *Dll* limb element from 294 to 731 (*Dll304Min*). Motif A, CACAATGC, is repeated twice in conserved positions, with a third instance of CACAAAGC nearby. Motif B, TTTGTT, is repeated twice within *Dll304Min* and once in the extended sequence, and is located within 5 bp of a Hox-like CAATTATG site, suggesting that

Motif B may be a cofactor that cooperates with a Hox protein to confer its regulatory effect. Mutating either site in the context of *Dll304Min* almost completely abolishes limb expression (Figure 3f, 3g). These mutations do not overlap known repressor binding sites (Gebelein *et al.*, 2004; Vachon *et al.*, 1992) or match known transcription factor binding sites, suggesting that motif A and motif B are bound by unidentified activators to drive *Dll* limb expression.

Dll304 is not identifiable in non-*Drosophila* insects based on sequence similarity

Distal-less is expressed in the developing limb in all tested arthropods (Panganiban and Rubenstein, 2002). To attempt to determine whether common *cis*-regulatory logic is used to control *Dll* limb expression in insects, I searched for homologs to *DmDll304* in *Anastrepha ludens, Musca domestica,* and *Anopheles gambiae.*

Based on conservation between the coding region of exon 1 of *D. melanogaster Dll* and publicly available *A. gambiae* genomic sequence, I designed degenerate primers accommodating all possible codons for the conserved amino acid sequence. Using these primers, I PCR-cloned *Dll* exon 1 from *A. ludens* and *M. domestica. Cis*-regulatory elements are rarely linearly conserved between *D. melanogaster* and other non-drosophilid flies*,* even if still functionally similar (Wratten *et al.*, 2006; Xiong and Jacobs-Lorena, 1995). However, since I cloned *Dll304* from *Scaptodrosophila lebanonensis,* an outgroup to all *Drosophila,* simply by degenerate PCR, I suspected that other species that diverged between muscatids and drosophilids, such as Tephritid fruit flies, might show linear conservation of some *Dll304* regulatory sequences. I designed inverse PCR primers to *A. ludens Dll* exon 1, and “walked” upstream of *Dll* to attempt to identify alignable regions to the *DmDll* locus. Despite sequencing over 12 kb of upstream sequence, I were unable to identify any homologous regulatory sequences. Additionally, no obvious clusters of motifs identified from bioinformatic analysis of *DmDll304* were found in this upstream region. I may not have sequenced enough upstream sequence to reach *Dll304*, but no regulatory elements located in the more proximal upstream *Dll* region are conserved either.

To attempt to clone *M. domestica Dll* flanking sequence, I used *MdDll* exon 1 as a probe to screen a Lambda phage *M. domestica* genomic library. Several attempts failed to detect any positive colonies. Estimates for *M. domestica* genome size vary from 295 to 950 megabases (Bier and Müller, 1969; Gao and Scott, 2006), so it is unclear whether our inability to detect *MdDll* in the genomic library was likely because of low genomic coverage.

We also attempted to identify *Dll304* from *A. gambiae* by looking for clusters of Hox sites and novel motifs that I identified in *DmDll304* (Figure 4a). Since no one region upstream of *AgDll* contained obviously clustered sites, I tested multiple segments covering the majority of the upstream region. *AgDllPt1* caused expression segmental stripes in *D. melanogaster* embryos (Figure 4b), and *AgDllpt2* drove weak expression along the ventral midline, but no tested fragments induced *Dll*-like expression (Figure 4b-4e).

Discussion

A frequent debate within the evolutionary/developmental biology (evo/devo) community is whether *cis*-regulatory or protein change is the major driving force to macro-evolution (Rodriguez-Trelles *et al.*, 2003). Protein evolution can be inferred between fairly distant relatives because of the constraints placed on how protein-coding DNA can change without rendering the encoded protein functionless. *Cis*-regulatory evolution is more difficult to determine over large timescales, as regulatory sequences change so quickly that it is impossible to recognize homologous sequences beyond fairly closely related species. Even when studying genes that are expressed in homologous tissues in diverged species, it is unclear whether the same regulatory logic is used to elicit this common expression, or whether Developmental System Drift has occurred (True and Haag, 2001). Since *Distal-less* is expressed in developing limbs in an incredible array of animals, it should be an optimal paradigm to test whether *cis*-regulatory logic is conserved, albeit in unalignable sequences, to drive conserved expression in developing limbs.

Dll cis-regulation in *Drosophila* species

In cases where homologous regulatory sequences can be identified by linear alignment, exogenous homologs usually recapitulate the function of the endogenous regulatory element (at least in published accounts). I found this to be true for the *Dll* embryonic limb regulatory element *Dll304,* as the *D. virilis* homolog drives reporter expression that is grossly similar to *D. melanogaster Dll304* expression. This confirms that the *D. virilis* homolog contains all motifs that are in *D. melanogaster* to regulate limb primordia expression, and most likely these motifs are contained within linearly conserved sequences between *D. melanogaster* and *D. virilis*. I was also able to clone *Dll304* from the more distantly related fly *S. lebanonensis*, and additional sequence analysis revealed extensive linear sequence conservation with *D. melanogaster*.

Activation of *Dll* transcription in embryonic limb primordia

Functional regulatory elements are composed of binding sites for one or more transcription factors that either promote or repress transcription of associated genes when bound. Multiple binding sites for a single regulator are often found within a single element, possibly to ensure some redundancy, or to "tune" transcriptional output depending on the number of sites. Techniques for *de novo* identification of motifs can take advantage of both this and the tendency for required binding sites to be conserved between species.

By limiting my search for repeated motifs to those that are conserved in identified *Drosophila Dll304* homologs, I greatly reduced the statistical "noise" introduced by the frequency that random sequence appears similar to a real motif. I identified several motifs that are repeated in linearly conserved positions in *Drosophila Dll304* homologs, strongly suggesting that they are binding sites for major regulators of *Dll* limb expression. Indeed, both novel motifs that I tested (Motif A: CACAATGC; motif B: TTTGTT) are required for activation. Since the other identified motifs lie outside of the minimal *Dll* limb element, I assume that a second regulatory element that controls an independent aspect of *Dll* expression is composed largely of these other identified conserved/repeated motifs.

A cluster of sequences in the 3' end of *Dll304Min* have been demonstrated to be required for repression of *Dll* expression in ventral epidermis of abdominal segments. These sites are presumably bound by UBX/ABD-A in complexes with EXD, HTH, EN, and SLP (Vachon *et al.*, 1992; Gebelein *et al.*, 2004) to repress transcription. In contrast, no published reports have identified any sequences required for activation, but unpublished deletion analyses demonstrated that removing the first 100bp of *Dll304Min* eliminates limb-specific expression (D. McKay, personal communication).

Genetic evidence may provide some clues to the function of required activation sequences in *Dll304Min.* Both the *wingless* (*wg*) pathway (Cohen *et al.*, 1993; Kubota *et al.*, 2003) and the transcription factor encoded by *buttonhead* (*btd*) (Estella *et al.*, 2003) are required for activation of *Dll* in thoracic limb spots, but no molecular evidence supports a direct role for either BTD protein or Pangolin (PAN), the *Drosophila* TCF/LEF-1 ortholog that transduces the *wg* signal to control transcription.

Several sequences that match BTD (GGGCGK) or PAN (BCTTTG) core consensus binding sites can be found in *Dll304Min*, including both a BTD and a PAN site within the first 100bp of *Dll304Min*. One of the required, conserved instances of Motif A (CACAATGC) is also within this region. Perhaps targeted mutations of each of these sites would reveal the required motif in this 100bp region. Motif A does resemble a loose match to TCF/Pangolin binding sites, and motif B is very closely associated with Hox-like sequences in all three locations in the *Dll304* region.

Dll cis-regulatory elements in other *Diptera*

Since *Dll* is expressed in limb primordia in all tested arthropods, we attempted to identify a functional regulatory element from near the *Dll* locus from several insects that diverged between 100 and 260 million years ago from *Drosophila.* We were unable to identify any sequences upstream of *Anastrepha Dll* that are linearly conserved to *Drosophila Dll*, suggesting that this evolutionary distance is generally too great to expect linear conservation of regulatory sequences, even for an essential gene with conserved expression. A more robust survey attempting to identify developmental *cis*-regulatory elements in *S. lebanonensis* and tephritids like *A. ludens* or *C. capitata* would hopefully reveal an outer bound of extensive linear sequence conservation. Sequencing a species just within this outer bound (presumably *S. lebanonensis*) would serve genome-scale *in silico* searches for functional *cis*-regulatory sequences and required motifs, since only the most essential regulatory sequences would be conserved. Only in the rarest cases is linear conservation observed between *Drosophila* and distant relatives, such as blowflies (Gibert and Simpson, 2003), or even individual obviously similar motifs in even more distant species (Erives and Levine, 2004; Rebeiz *et al.*, 2005; Wratten *et al.*, 2006; Xiong and Jacobs-Lorena, 1995).

The inability of any tested *A. gambiae Dll* upstream sequence to recapitulate *Drosophila* embryonic *Dll* expression is possibly due to inadvertent "splitting" of a functional regulatory element between multiple tested construct fragments, the exclusion of essential sequences from consideration because of improper annotation, or the relocation of embryonic limb regulatory elements to *A. gambiae Dll* introns. It is entirely possible, however, that a different limb *cis*-regulatory code evolved in *A. gambiae,* using binding sites to transcriptional regulators that are expressed in different patterns in *A. gambiae* and *D. melanogaster.*

Re-annotation of the released *A. gambiae* genomic sequence indicated a region upstream of the four tested elements, which had previously been annotated as the first exon of a separate gene, is in fact intergenic sequence. It is possible that the *A. gambiae Dll304* homolog exists in this region, with an instance of Motif A located 800bp from an instance of Motif B near a Hox-like site in *AgDllPt1*.

Unlike *DmDll304*, which has very tight clustering of Motifs A and B, no similar clusters were observed in *A. gambiae, A. mellifera*, or *T. castaneum* (data not shown). Again, artificial definitions of motif consensus sequences may be both eliminating true binding site matches while identifying spurious matches. Alternatively, Developmental System Drift may have changed one or more of the activating regulatory signals, in which case these clustered sequences would quickly disappear during evolution. A comprehensive test of the non-coding sequences of *Dll* from several distant insect species, both in *D. melanogaster* and the native species, along with *DmDll304* tested in those animals, using a polyspecific transformation system such as PiggyBac (Grossman *et al.*, 2001; Handler and Harrell, 1999) would differentiate between these possibilities.

Materials and Methods

Insect stocks and genomic DNA: *D. melanogaster* strain *w^{1118}* was used for germline transformation (Rubin and Spradling, 1982; Spradling and Rubin, 1982), *in situ* hybridization, and source for genomic DNA. Fly stocks for *D. pseudoobscura, D. virilis, D. immigrans, D. hydei,* and *Scaptodrosophila lebanonensis* were supplied by the Tucson *Drosophila* Stock Center (Tucson, Arizona). *D. virilis* λ phage library was a generated by Thomas Kaufman and supplied by Par Towb. *M. domestica* λ phage library was a kind gift from Jeff Scott (Cornell University, Ithaca, New York). *Anastrepha ludens* adults were kind gifts from Kevin Hoffman at the Medfly Exclusion Program (Los Angeles, CA). *Anopheles gambiae* genomic DNA and embryos were supplied by ATCC. Genomic DNA was prepared using standard procedures.

λ phage library screening: *D. virilis* and *M. domestica* libraries were screened on nitrocellulose membranes with radioactive probes of *D. melanogaster* 877 bp *Dll304* (*Dv* and *Md*), or a fragment of *M. domestica Dll* exon 1, labeled by nick translation. Hybridization and washes were carried out at 37°C. Positive *D. virilis* clones were amplified and purified using Qiagen Lambda Maxi Kit. A positive clone was cut with HindIII and probed with *DmDll304* using Southern Hybridization. HindIII-cut clone DNA was subsequently cloned into pBluescript KS+ HindIII sites, PCR screened for inserts of ~1.8kb, and sequenced to confirm. Additional flanking sequences were obtained by inverse PCR on genomic DNA.

PCR and Inverse PCR: PCR primers generated by IDT (Coraville, Iowa) were used for either classical or inverse PCR, using a standard Touchdown PCR protocol, on genomic or phage DNA. Nested primers were sometimes used for Touchdown PCR to minimize spurious products. Inverse PCR protocol was based on BDGP protocol (www.fruitfly.org). Primer sequences available upon request.

Germline Transformation of *Drosophila*: *w*^{1118} embryos were transformed using standard protocols with pH-Stinger expressing either GFP or DsRed (Barolo *et al.*, 2000; Barolo *et al.*, 2004).

Construct boundaries and sequence alterations:

DmDll304: 877 bp fragment from digestion with EcoRI and SspI.

DvDll304: 1199 bp fragment based on sequence conservation with *Dll304*, from AAGCTTATTTTAGGAATGTA to AAATAGGATTTGCGT.

Dll304Min: is composed of nucleotides 294 to 731 of *DmDll304*.

Dll304Min-MotifA: changed AGG***GTGTC***AGCCAG***GTGTC***TGC and CCA***GTGTC***TGC

Dll304Min-MotifB: changed AGC***TGACT***AAG and GCA***TGACT***ACC

AgDllPt1: CGATTGTCAAAG to TAACGTCCTAC

AgDllPt2SubA: CTTACCGGGTGATG to CAAAGGCAGG

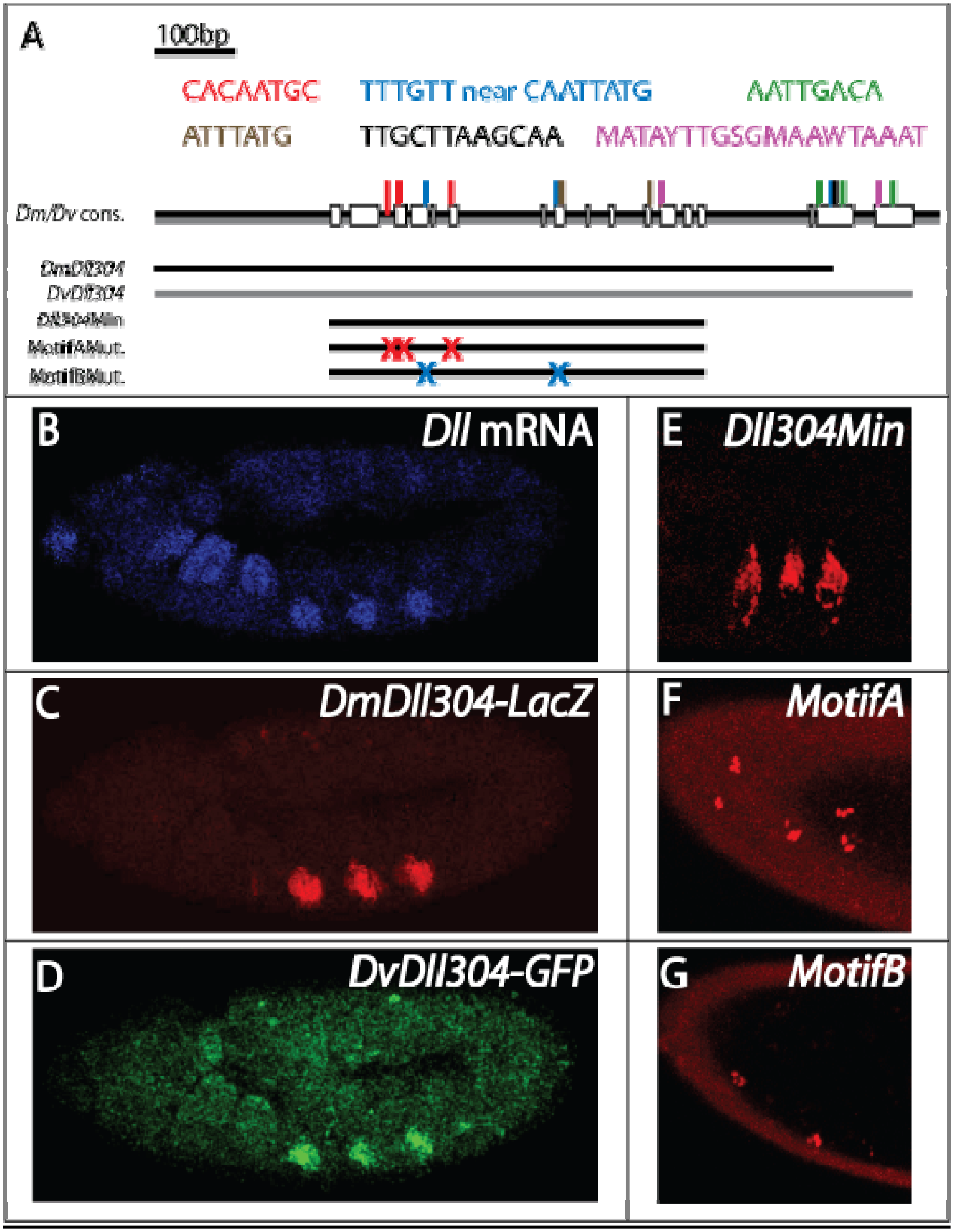
AgDllPt2SubB: GTGGTTGAGAC to GCGAACCGTC

AgDllPt2SubC: CATAAACCAGCG to CCTTACAATTCA

Multiplex Fluorescent *In Situ* Hybridization: Probes were generated from full-length clones of LacZ and GFP, and a partial *Dll* cDNA from 5' start to an EcoRI site, into which Digoxigenin, Biotin, and Di-nitrophenol-labeled UTP, respectively, were incorporated. Hybridization protocol was as described by Dave Kosman's MFISH protocol (Kosman, 2004).

Sequence Alignments: Sequences cloned by molecular biological techniques or identified Discontiguous MegaBLAST of NCBI Trace archives were trimmed to approximate common boundaries and aligned by T-Coffee (Notredame et al., 2000). Alignments were then adjusted based on evolutionary proximity based on Lalign (http://www.ch.embnet.org/software/LALIGN_form.html) pairwise alignments of small sections of poorly aligned sequences.

Figure 3. Multiple repeated motifs that are conserved in *Drosophila Dll304* are required for activation. (A) Schematic of *Dll304* region, tested constructs and conserved/repeated motifs. Top: White rectangles indicate blocks of significant conservation between *D. melanogaster* and *D. virilis*. Colored lines indicate positions of matching motif instances. Bottom: tested elements based on *Dll304*. X marks mutated sites in relevant constructs. **(B-D)** Triple fluorescent *in situ* for (**B**) *Dll* transcript, (**C**) *LacZ* under the control of *DmDll304*, and (**D**) *GFP* transcript under the control of *DvDll304*. (**E)** Closeup of embryonic thoracic segments in stage 13 embryo containing a construct with *Dll304min* driving DsRed in thoracic limb spots. **(F,G)** Mutating sites matching either (**F**) Motif A (CACAATGC) or (**G**) Motif B (TTTGTT) in *Dll304Min* almost completely abolishes DsRed expression in thoracic spots.


A
100bp
CACAATGC
TTTGTT near CAATTATG
AATTGACA
ATTTATG
TTGCTTAAGCAA
MATAYTTGSGMAAWTAAAT
Dm/Dv cons.
DmDll304
DvDll304
Dll304Min
MotifAMut.
MotifBMut.
B
Dll mRNA
C
DmDll304-LacZ
D
DvDll304-GFP
E
Dll304Min
F
MotifA
G
MotifB

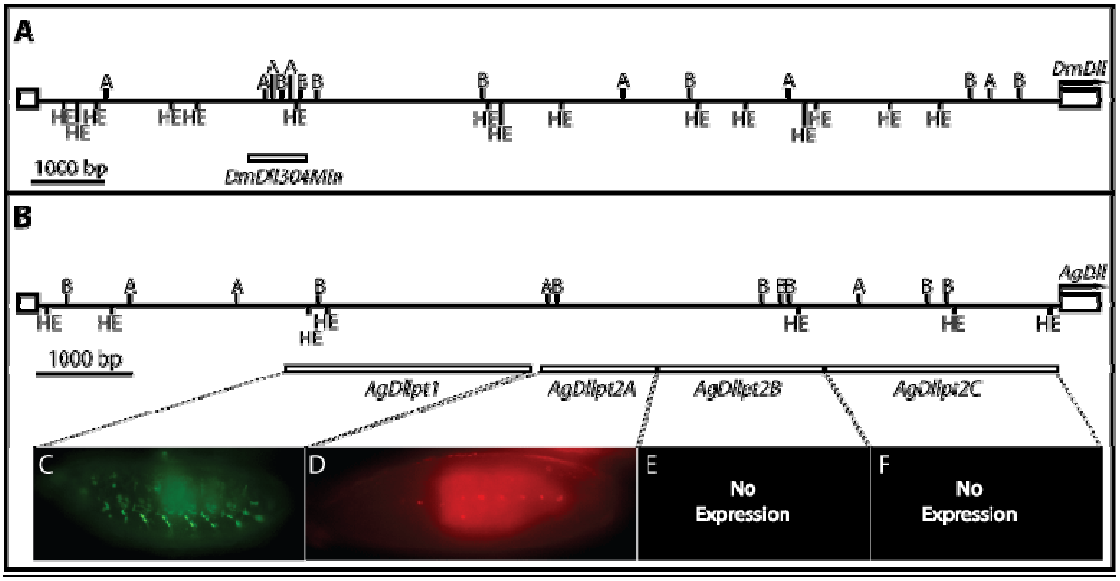


Figure 4. ***A. gambiae Dll*** **upstream sequences do not drive limb expression in *D. melanogaster.*** **(A)** Schematic of *D. melanogaster Dll* upstream region. Matches to Motif A (CACAAWGC) and Motif B (TTTGYT within 10bp of ATTA), and matches to *in vivo* Hox-EXD binding sites are indicated by A, B, and HE, respectively. Bounds of *Dll304Min* is indicated below. **(B)** Schematic of *A. gambiae Dll* upstream region. Tested genomic fragments from *AgDll* upstream is indicated below. (**C)** Weak ventro-lateral segmental expression was observed in embryos containing *AgDllPt1*-GFP reporters. (**D)** Weak ventral spots of DsRed expression were observed in *AgDllPt2A* embryos. (**E,F)** No DsRed expression was observed in embryos containing *AgDllPt2B orAgDllPt2C*-DsRed reporters.

Chapter III

Common *Cis*-Regulatory Logic of the *Drosophila* Wound Response

Introduction

All organisms, regardless of size or lifespan, are in constant danger of being wounded. If not repaired, these wounds are inevitably fatal, due to infection, nutrient loss, or simply desiccation. To combat this, intricate systems have evolved to heal injuries and protect against invaders. In recent years, it has become apparent that many aspects of these responses are common between widely diverged groups, such as between vertebrates and invertebrates, suggesting that these responses may have even existed in the bilatarian ancestor to these animals. Included in these conserved responses are aspects of the innate immune response (Hoffmann and Reichhart, 2002), inflammatory response (Bokoch, 2005; Stramer *et al.*, 2005), clotting (Karlsson *et al.*, 2004), and re-epithelialization (Martin and Parkhurst, 2004).

The outermost layer of the mammalian skin barrier is the stratum corneum, a constantly regenerated layer of cross-linked skin cells, proteins, and lipids. This layer prevents water loss, resists mechanical and chemical penetration, and microbial invasion (Alibardi and Kwang, 2006). The analogous insect structure to the stratum corneum is the cuticle, comprised of cross-linked chitin, proteins, and lipids secreted by the underlying epidermis. First secreted in late embryogenesis, cuticle serves as a hard barrier against injury and desiccation.

Aseptic wound healing mechanisms are remarkably well-conserved between vertebrates and invertebrates. Both *Drosophila* and amniote embryos heal wounds with similar mechanisms: An actin cable surrounds the wound (Martin and Lewis, 1992; McCluskey and Martin, 1995; Wood *et al.*, 2002), under the control of Rho

GTPases (Brock *et al.*, 1996; Wood *et al.*, 2002), closing the wound via a "purse-string" mechanism. This mechanism is reminiscent of *Drosophila* embryonic dorsal closure(Young *et al.*, 1993) and *C. elegans* ventral enclosure (Williams-Masson *et al.*, 1997). Even the mechanism by which larval and adult insects heal wounds is superficially analogous to mammalian adult wound healing. In *Drosophila,* wounds are quickly sealed by a plug of cell debris, and the plug is rapidly cross-linked, preventing acute water loss (Jiravanichpaisal *et al.*, 2006). Epidermal cells then move together under the plug to reform a continuous epithelium, and secrete new cuticle to seal the hole in the exoskeleton, leaving a scar as evidence of the injury (Galko and Krasnow, 2004; Ramet *et al.*, 2002).

The major components of insect cuticle, chitin and cuticle proteins (Andersen *et al.*, 1995), are cross-linked by highly reactive quinones, which are derived from enzymatically-processed tyrosine. These reactions occur both during development and during the cuticle regeneration step of wound healing. The extent of cross-linking regulates structural strength (Vincent and Wegst, 2004). Two enzymes in the quinone-generation pathway, encoded by the genes *Dopa decarboxylase (Ddc)* and *pale (ple),* are transcribed at extremely high levels around late embryonic wounds, presumably to increase the pools of precursors to quinones that crosslink newly-secreted cuticle to repair the rupture. C*is*-regulatory elements controlling this non-infectious wound-induced transcription were identified (Mace *et al.*, 2005), which we have dubbed Wound Response Elements (WREs).

Bioinformatic analyses of phylogenetically conserved WRE sequences revealed several motifs within the *Ddc* WRE (Mace *et al.*, 2005), including motifs that match AP-1 (FOS/JUN heterodimer) consensus binding sites and a Grainy head (GRH) consensus binding site that had been previously shown to be required for larval CNS *Ddc* expression (Bray *et al.*, 1988). Both of these motifs are required for *Ddc* WRE function (Mace *et al.*, 2005).

The *Ddc* WRE was also non-functional in *grh*^{IM} mutants, but maintained wound response activity in tested *fos* (*kay*^{1}) and *jun* (*Jra*^{IA109}) mutant embryos. While the protein made from the tested *Jra*^{IA109}allele is truncated before the dimerization and DNA binding domains and is thus presumably non-functional, *kay*^{1} only affects one of four isoforms, leaving open the possibility that another FOS isoform is regulating the *Ddc* WRE, either as a homodimer or a heterodimer with another bZIP protein.

Based on results from the *Ddc* WRE, we identified two WREs for *pale (ple)* by searching for conserved motifs matching GRH (ACYNGTT) and AP-1/FOS/bZIP (AFB) (TGANTCA) consensus sites. The identification of multiple WREs by searches for clusters of AFB and GRH consensus sites suggested that a common regulatory mechanism may activate multiple wound response. This mechanism could be quite ancient, as both JUN, FOS and GRH proteins are involved in the mammalian wound response (Ting et al., ; Yates and Rayner, 2002).

To better characterize the transcriptional wound response in *Drosophila* embryos, we have further dissected sequence requirements of the *Ddc* and *ple* WREs. We have determined the minimal sequence requirements for the *Ddc* WRE, and identified several new motifs that affect *Ddc* WRE function. We confirmed that the motifs used to identify the distal *ple* WRE are required for the embryonic wound response. By searching for clusters of conserved sites matching AP-1 and GRH consensus sites, we identified two new WREs for the genes *krotzkopf verkehrt (kkv)* and *misshapen (msn).* Surprisingly, these WREs differ in their ability to function in *grh*^{IM} or *kay*^{sro} mutants, as well as their activity in larvae and adults. Given the diverse nature of the proteins encoded by these wound response genes and the complexity of the epidermal response, we expect that this wound response *cis*-regulatory code is likely to be quite prevalent in the genome of *Drosophila*, and possibly in vertebrates.

Results

Minimal *Ddc* Wound Response Element sequence requirements

We previously demonstrated that *Dopa decarboxylase* (*Ddc*) is transcribed around epidermal wounds, and identified upstream sequences that are sufficient to recapitulate this response (Mace *et al.*, 2005). A fragment from -1.4 kilobases (kb) to transcription start (*Ddc -1.4*) is a functional Wound Response Element (WRE) (Mace *et al.*, 2005). Two subfragments, containing sequences from -1.4 kb to -.38 kb (*Ddc*Δ*-380*) (K. Mace, unpublished) or from -.47 kb to transcription start (*Ddc-.47*} (Mace et al., 2005), both function as WREs. The overlapping 90 bp from -.47 kb to -.38 kb is not sufficient as a WRE, as mutating a GRH site outside of this region in *Ddc -.47* abolishes WRE function (Mace et al., 2005). This GRH is not in *Ddc Del. -380*, but a second GRH-like site within this element is most likely substituting as the required GRH binding site.

The 90 base pair overlap from -.47 to -.38 kb shared between the *Ddc Del. -380* and *Ddc -.47* WREs includes 44 bp that are conserved through *D. virilis,* called Conserved Region 1 (CR1). The *D. virilis* homolog is functional as a WRE in *D. melanogaster* (Mace et al., 2005), strongly suggesting that the conserved sequences contribute to WRE function. To test whether these conserved sequences are required for WRE function of *DdcΔ-380* or if two separate WREs exist from -1.4 to -.47 kb and -.47 to -0.0, I deleted CR1 from the full *Ddc-1.4* WRE and *DdcΔ-380*, generating *Ddc-1.4ΔWRE* and *Ddc-1.4to-.47.* I cloned these DNA fragments into the DsRed H-Stinger P-element vector (Barolo *et al.*, 2004), and injected the constructs into *D.*

melanogaster embryos, testing multiple lines of resulting transformants for DsRed fluorescent protein expression around epidermal wounds caused by glass microinjection needles. Neither element was able to drive DsRed expression around wounds, except for a few rare cases where very weak expression was observed, demonstrating that CR1 is required for any *Ddc* WRE function (Fig. 6d,e).

The distal conserved region, CR1, contains a sequence matching the consensus binding sites for AP-1 (TGAcTCA) (Pollock and Treisman, 1990). However, we previously found that the *Ddc-1.4* WRE is still fully functional in homozygotes for an amorphic *jra* allele (Mace et al., 2005), strongly suggesting that this site is not in fact a canonical AP-1 site, bound by JUN-FOS heterodimers. *Drosophila* FOS, unlike mammalian FOS orthologs, is able to homodimerize *in vitro* and bind to TGANTCA sequences (Perkins *et al.*, 1988), as are heterodimers of FOS with CREB (Eresh *et al.*, 1997; Masquilier *et al.*, 1992), and other predicted dFOS/bZIP heterodimers (Fassler *et al.*, 2002) are quite likely to also recognize this site, at least *in vitro*. Thus, we refer to sites matching the TGANTCA consensus as AP-1/FOS/bZIP (AFB) sites, to avoid implying that JUN/FOS heterodimers are binding these sites. Immediately adjacent to the conserved AFB site is a set of three clustered sites that resemble ETS family binding sites (cmGGAWgy) (Sharrocks *et al.*, 1997).

The region from -433 to -72 bp is very poorly conserved, with only small segments alignable between *D. melanogaster* and *D. pseudoobscura*, with no detectable linear conservation to *D. virilis.* A second Conserved Region (CR2), from -71 to -58, was originally identified based on conservation with *D. virilis* (Bray and

Hirsh, 1986), and contains a perfectly conserved GRH consensus binding site (ACYgGTT) (Venkatesan *et al.*, 2003) overlapping a Tramtrack (TTK) consensus binding site (GGTCCTGC) (Read *et al.*, 1990). The final conserved block matches the TATA motif in the proximal promoter.

To test whether any required DNA elements are located within the region between CR1 and CR2, I generated three overlapping ~125 bp deletions from the *Ddc -.47* WRE. All three deletions still function as WREs, although the first deletion (*Ddc-.47Δ1*), which removes two ETS consensus site matches, is somewhat weaker than the wild-type element (Fig. 6f). The other two deletions (*Ddc-.47Δ2, Ddc-.47Δ3*) show no difference in timing or intensity of DsRed expression compared to the wild-type element (Fig. 6g,h). To confirm that no redundant sequences within the 362 bp region contribute to *Ddc -.47* WRE function, we deleted the entire unconserved fragment from the *Ddc -.47* WRE (*Ddc-.47Δ123*), leaving a 117 bp fragment. This element also functions as a WRE, but while the breadth of the activation from the wound site is comparable to *Ddc-.47,* the number of nuclei within this radius is noticeably reduced compared to wild-type *Ddc -.47* (Fig. 6i). Nonetheless, the ability of *Ddc-.47Δ123* to act as a WRE confirms that no sequences between CR1 and CR2 are required for *Ddc* wound-induced activation.

Identification of Sites Required for Maximal *Ddc* WRE Function

Bioinformatic searches for known transcription factor binding sites within *Ddc -.47* revealed several matches to consensus binding sites for putative regulators (Fig.

7a). To determine whether any motifs within *Ddc -.47* that match binding sites for known transcription factors affect *Ddc -.47* expression, we altered these motifs in an attempt to make them unrecognizable to those transcription factors. First, we tested sites in the CR1 region, including sites matching AFB and ETS family consensus binding sties.

In addition to the perfectly conserved match to the AFB consensus in CR1, we found a second match that lies within the region deleted by *Ddc -.47Δ2*. This second site is conserved through *D. persimilis*, but not *D. pseudoobscura*, its sister species. Mutating both sites eliminated WRE function in DsRed reporters, strongly suggesting that the AFB consensus site is the sequence within CR1 that is required for *Ddc* WRE function (Fig. 7d).

Three sites clustered immediately promoter-proximal to CR1 are reminiscent of ETS family binding sites. Mutating all three clustered matches, none of which are linearly conserved in *D. virilis*, noticeably weakened, but did not eliminate WRE function (Fig. 7e,f). Two of these sites are deleted in *Ddc-.47Δ1* and *Ddc-.47Δ123* (Figs. 6a,6f,6i, & 7a), so perhaps the missing ETS-like motifs lead to the reduction of WRE function in those deletions.

CR2 consists of 16 bp perfectly conserved in all sequenced drosophilids, and includes a site matching the GRH consensus sequence that is required for *Ddc* CNS expression (Bray *et al.*, 1988; Scholnick *et al.*, 1986), as well as *Ddc -.47* WRE function (Mace et al., 2005). Overlapping this site is a perfectly conserved sequence that matches the TTK consensus site (Read *et al.*, 1990). To attempt to avoid altering

GRH binding, we altered the putative TTK sequence at two nucleotides adjacent to the GRH site consensus site. We observed a significant reduction in the *Ddc*TTK WRE's activation (Fig. 7g, h), but it is possible that we inadvertently altered an extended GRH binding site that is not reflected in the published consensus.

ple Distal WRE requires conserved sites matching GRH and AFB consensus sites

We previously identified two WREs upstream of *pale (ple)* by searching for conserved clusters of AFB and GRH-like sites (Mace et al., 2005). We refined the boundaries of the distal *ple* WRE within the 3 kb fragment, identifying a 687 bp fragment that is indistinguishable as a WRE from the 3 kb element (Fig. 8C). Bioinformatic analysis of the conserved sequences revealed several potential binding sites for other transcription factors in addition to the AFB and GRH consensus sites, including sites matching consensus binding sites for CREB homodimer (TGACGTMA) (Benbrook and Jones, 1994), an Extradenticle half-site (EXD, TGAT) (Neuteboom and Murre, 1997; van Dijk and Murre, 1994; van Dijk *et al.*, 1993), and Hox family monomer transcription factors (ATTA) (Ekker *et al.*, 1991; Pearson *et al.*, 2005; Pellerin *et al.*, 1994).

To determine which, if any, of these sites contribute to the element's WRE function, we mutated all sites matching consensus sequences for these transcription factors. In addition to canonical matches to AFB, CREB, and EXD consensus sites, a conserved sequence, TGATTGAC, was also found that resembles these consensus sequences. To ensure that we did not leave functional sites intact, we mutated this site in addition to the canonical AFB, CREB, or EXD site matches, in appropriate constructs (Fig. 8B).

Mutating the canonical AFB site along with the AFB/CREB/EXD-like site abolished *pleSubBMin* WRE function (Fig. 8D). Similarly, mutating the GRH consensus site almost completely abolished *pleSubBMin* WRE function (Fig. 8E). Thus, the two motifs identified by *Ddc* WRE dissection and used to identify the *ple* WREs are required for *ple* wound response.

In contrast, mutating the AFB/CREB/EXD site along with either the canonical CREB-like or EXD-like sites had no detectable effect on *pleSubBMin* WRE function (Fig. 8 F,G). This suggests that these other sites regulate other aspects of *ple* expression, but not the wound response. Similarly, Mutating all fourteen Hox-like sites, twelve of which are conserved through *D. virilis*, does not noticeably affect *ple* WRE function (Fig. 8H). The clustering of required AFB and GRH consensus motifs within the 5' half of *pleSubBMin* suggest that the functional *ple* WRE is located in this small sub-fragment, and uses similar *cis*-regulatory logic to the *Ddc* WRE.

Identification of novel WREs by clustering of AFB and GRH consensus sites

To attempt to identify new WREs by searching *in silico* for conserved clusters of sequences matching AFB and GRH consensus sites, we searched in loci for two candidate genes that we suspected would be up-regulated during the wound healing process. *krotzkopf verkehrt (kkv)* encodes chitin synthase (Ostrowski *et al.*, 2002), which is required for the final step in synthesis of chitin, a major component of

Drosophila exo- and endocuticle (Merzendorfer and Zimoch, 2003). *misshapen (msn)* encodes the MAPKKKK upstream of the Jun Kinase encoded by *basket* (Su *et al.*, 1998), which phosphorylates the AP-1 proteins dJUN and dFOS. Previous studies have demonstrated that a LacZ enhancer trap insertion in the *msn* locus (Spradling *et al.*, 1999) is activated near larval (Galko and Krasnow, 2004) and adult (Ramet *et al.*, 2002) epidermal wounds. Using Multiplex Fluorescent *In Situ* Hybridization (MFISH) (Kosman *et al.*, 2004), we detected rapid transcriptional activation of *kkv* and *msn,* as well as *Ddc* and *ple,* near epidermal wounds in *Drosophila* embryos (Fig. 9). No increased expression was observed in embryos that were wounded and immediately fixed or bisected after fixation, demonstrating that the observed up-regulation was not due to accessibility artifacts (data not shown). All four genes were detected within 30' post-wounding, suggesting that all may be regulated by similar *cis*-regulatory codes.

To identify putative WREs regulating *kkv* and *msn* transcription, we surveyed the respective loci for clusters of AFB and GRH consensus site matches, then checked whether identified clustered sites were conserved in *D. pseudoobscura* and *D. virilis*. Within the first intron of *kkv,* we identified a cluster of conserved AFB and GRH consensus sites. When tested in a reporter construct, a 2.2 kb fragment containing these sites functioned as a WRE (*kkv1*, Fig. 10A,D)). We also identified the *msn* WRE as a 2.3 kb fragment containing 3 GRH and 1 AFB consensus sites, located 8.7 kb downstream of transcription start, in the third intron (*msn1.2*, Fig. 10B,C). We subsequently identified a functional 1.2 kb subfragment of the *msn* WRE (*msnSubB)* containing all GRH and AFB consensus sites.

To confirm that the sites matching AFB and GRH consensus motifs used to identify the *kkv1* WRE are required for activation, we altered all sites resembling either AFB or GRH consensus sites in *kkv1* (Figure 11A). Surprisingly, the *kkv1* WRE requires neither AFB (Fig. 11C) nor GRH (Fig. 11D) consensus sites for wound-dependent activation.

Multiple *Trans*-Regulators activate WREs through AFB and GRH consensus sites

All identified WREs contain at least one conserved sequence matching GRH and AFB consensus sequences, and altering all matches to either set of sites in the *ple* and *Ddc* WREs essentially eliminates activation in response to wounding. To determine whether these identified motifs are indeed bound by the presumed transcriptional regulators to activate wound transcription, we tested for *in vitro* interactions and genetic requirements of GRH and AP-1 proteins.

In vitro translated dFOS/dJUN heterodimers able to bind oligos containing the conserved *Ddc* AP-1 consensus site (data not shown), and all identified AFB consensus matches in the other WREs do not differ significantly from the *Ddc* site or the AP-1 consensus and match sequences previously shown to be bound by dFOS, AP-1, and CREB proteins (Perkins *et al.*, 1988; Pollock *et al.*, 1990; Zhang *et al.*, 1990).

We previously tested the *Ddc -1.4* WRE in zygotic mutants for *bsk*, *jra,* and *kay* (Mace *et al.*, 2005). We saw no reduction in WRE function, and observed activation at the "wound" of the failed dorsal closure phenotype of these mutants. These data apparently conflict with the presence of conserved AFB consensus sites in all tested WREs. We had previously eliminated *jra* as the factor binding the *Ddc* AFB sites (Mace *et al.*, 2005), but the tested *kay*^{1} mutant did not eliminate all isoforms of FOS. To attempt to resolve this conflict, we tested WREs in a *shroud* (*sro*) mutant, which was recently identified as a mutation in an exon of *kay* that is highly expressed in late embryonic epidermis based on a P-element insertion upstream of a previously-unknown exon of one FOS isoform (Giesen *et al.*, 2003). To test whether *kay*^{sro} is required for WRE function, we introduced *pleSubBMin, kkv1,* and *msn1.2* WREs into an EMS-induced *kay*^{sro} mutant line background. Surprisingly, the *pleSubBMin* and *msn1.2* WREs are not activated after wounding in *kay*^{sro} homozygotes. However, the *kkv1* is still activated at wound sites in *kay*^{sro} homozygotes (M. Juarez, unpublished observations). This suggests that an isoform of FOS that is affected in *kay*^{sro} mutants is required for *msn* and *ple* wound response, but not *kkv*.

Considerable published evidence establishes the requirement of GRH for activation of *Ddc* in developmental and wound-induced epidermal expression. In *grh*^{IM} mutants, *Ddc -1.4* is not activated at wounds (Mace *et al.*, 2005). This corresponds with the *cis*-requirement of a previously identified GRH binding site (Mace et al., 2005). The GRH consensus site matches in other identified WREs vary considerably in similarity to each other and relative to the optimal GRH binding site, AACCGGTT. The strongest GRH binding sites in WRES for *Ddc* (GACCGGTT), and *msn* (AACCGGTT) are well-conserved and strongly match the core optimal site.

The strongest *kkv1* GRH-like site (ACTGGTT) matches the weaker GRH consensus ACYGGTT. The minimal *ple* WRE, however, only contains one weakly conserved GRH-like site (ACTCGTTT) that matches the degenerate GRH consensus ACYNGTTT.

To test whether GRH can recognize the *ple* GRH-like site, I expressed a truncated form of GRH (Uv *et al.*, 1994) in *E. coli*, and used crude cell extracts in an Electrophoretic Mobility Shift Assay (Fried and Crothers, 1981) to test for binding to the oligos of sequences surrounding the strong GRH consensus site from the *Ddc* WRE and the weak site from the *ple* WRE. *E. coli* extract expressing GST did not bind either site, but extract containing GRH recognized both the *Ddc* and *ple* sites (Fig. 12). More *Ddc* GRH site probe was bound by GRH-BE compared to the *ple* GRH site probe, suggesting that this site is closer to the optimal GRH binding site. Mutating the *Ddc* GRH site in the same manner as the *Ddc -.47*GRH reporter construct abolished GRH binding, while mutating adjacent nucleotides that were changed in *Ddc -.47*TTK reduced GRH affinity. Mutating the *ple* GRH site strongly weakened, but did not fully eliminate GRH binding. These data confirm that the sites identified *in silico* as GRH consensus sites are recognized *in vitro* by GRH protein.

To test whether the *pleSubBMin, kkv1,* and *msn1.2* WREs require *grh,* I tested for WRE activation in zygotic *grh*^{IM} homozygotes and heterozygotes. Neither *kkv1* nor *pleSubBMin* activation is noticeably reduced (Fig. 13 a-d). In contrast, the *msn* WRE is substantially weaker in *grh*^{IM} homozygotes (compare Fig. 13 e,f), strongly suggesting that GRH regulates *msn* wound response through canonical GRH binding sites in the identified *msn1.2* WRE.

WRE Activation in Larvae and Adults

Prior to cuticle deposition in embryonic stage 16 (Campos-Ortega and Hartenstein 1997), wounds are healed by the "purse-string" mechanism (Wood *et al.* 2002), while wounds caused in older animals are healed by epidermal cell fusion and migration to close the wound, followed by cuticle synthesis (Galko and Krasnow, 2004). Prior to ~13 hours at 25°C, we are unable to detect activation of any WRE (I. Lidsky, unpublished observations), consistent with the hypothesis that the identified WREs control genes that are involved in larval-type wound healing mechanisms such as epidermal spreading and cuticle regeneration. Similarly, we do not see *Ddc* WRE function in early 1st instar larvae, but we do observe activation in late 1st instar larvae (42-48 hrs) and newly eclosed adults. Surprisingly, we do not see activation of the *pleSubBMin* or *kkv1* WREs around wounds induced in first-instar larvae or adults (I. Lidsky, unpublished observations). Considerable larval and adult epidermal expression is observed in unwounded animals containing *pleSubBMin* and *kkv1* reporters, which may obscure subtle wound-induced expression. It is also possible that multiple WREs for different developmental stages have evolved for these genes and are located outside of tested fragments, or that *ple* and *kkv* are not activated at larval or adult wounds.

Discussion

We have identified a set of motifs, GTGANTCA and ACYNGTT, that are linearly conserved in all tested drosophilids in at least four Wound Response Elements (WREs). These WREs activate a diverse set of genes in response to wounding, which are involved in epidermal migration and cuticle production and cross-linking. These motifs match consensus sequences for transcription factors that are known to be required for epidermal development and wound healing in both vertebrates and invertebrates, suggesting an ancient origin of this wound-dependent transcription mechanism. Surprisingly, the most likely candidate transcription factors for these sites are only required for subsets of the WREs, suggesting a complex regulatory system has evolved to regulate the wound response through this common set of binding sites.

Cis-Regulatory Motif Requirements of *Drosophila* Wound Response

Considerable research dissecting *Ddc* regulatory sequences upstream of the gene has revealed that different segments of upstream sequences regulate *Ddc* epidermal vs. CNS expression. CR1 is required for *Ddc* WRE function even in the context of the largest element, *Ddc -1.4*, despite the presence of several other sequences matching AFB consensus sites upstream and downstream of CR1. In contrast, while mutating the GRH site in CR2 abolishes WRE function of *Ddc -.47*, removing it completely in *DdcΔ-380* does not affect wound-induced activation. It is likely that a second site matching a GRH consensus site (Uv *et al.*, 1997) at -591 can substitute for the proximal site in its absence. Oddly, multimers of this distal GRH consensus site can drive epidermal expression, while multimers of the proximal GRH consensus site that is required for *Ddc-.47* WRE function drives CNS expression (Uv *et al.*, 1997).

Deletions of the entire 361 bp region between CR1 and CR2 only had a modest effect on the *Ddc* WRE. The sequences between CR2 and the presumed TATA motif in the proximal promoter (Bray and Hirsh, 1986) are not linearly conserved in all drosophilids, but degenerate motifs can be found that are common to all species. Nevertheless, it is likely that the AFB and GRH consensus site matches are the primary sites required for activation. The reduction of activation when the ETS consensus sites are altered and the reduced activity of *Ddc -.47Δ1* and *Ddc-.47Δ123* may be due to mutation or deletion of the same motif, but this site only contributes moderately to *Ddc* WRE function. The reduction seen in the *Ddc-.47*TTK WRE is likely reducing GRH binding to the adjacent site, as reduced GRH binding is seen to a probe containing this same site. TTK may still play a role through this site in GRH-dependent CNS expression of *Ddc*.

We previously identified two independent upstream regions of *ple* that drive wound-dependent expression by searching for AFB and GRH consensus sites that were conserved in *D. virilis* (Mace *et al.*, 2005). The distal WRE, which activates reporter expression almost as quickly as the *Ddc* WRE, contains several strong matches to both AFB and GRH consensus sequences. Progressive deletions from the ends of the original 3 kb element, using blocks of conservation as guides for endpoints, led us to the discovery of a 687 bp element that is sufficient to recapitulate both *ple's* wound response and anal pad expression. The two identified required WRE motifs in *pleSubBMin* are only separated by 7 bp in the 5' half of the element. The other tested motifs, all of which save a CREB-like site near the 5' end of *pleSubBMin*, are located in the 3' half in a highly conserved area and have no effect on wound-induced activation. All constructs with an altered version of a conserved ambiguous bZIP/EXD-like site alter anal pad expression, while altering all fourteen ATTA sites in the 3' half of *pleSubBMin* abolishes anal pad expression. This suggests that an even smaller element in the 5' half could be identified that only controls wound expression, while the rest of the element controls additional expression, including anal pad expression. These data also confirm that the *cis*-regulatory wound code regulating both *Ddc*'s and *ple*'s wound response is composed of a small set of conserved motifs matching AFB and GRH consensus sites.

Identification of novel WREs by clustering of Motifs

The wound healing process is complex, involving healing of both the epidermal and cuticular hole. This requires regulation of cell migration, enzyme synthesis, and secretion of cuticle components. The WREs that we have identified regulate genes necessary for cuticle synthesis, cuticle sclerotization, and epidermal migration.

Although all WREs identified contain conserved motifs matching AP-1 and GRH consensus sites, we were quite surprised to find that these sites are not required in the context of the *kkv1* WRE, as altering these sites in the same manner as the *Ddc* and *ple* WRE site mutations had no effect on wound activation of reporters *in vivo*. Consistent with this, neither tested subelement of *kkv1*, which both contained AP-1 and GRH consensus sites, had any wound response activity. It is possible that redundant mechanisms can compensate for altered AP-1 and GRH consensus sites in *kkv1*, or that a completely independent mechanism is at work and we happened upon the WRE by chance. Another putative WRE containing more tightly clustered conserved AP-1 and GRH consensus sites is located in the far 3' end of the same intron as *kkv1*, which may be another redundant WRE, similar to the situation we found with *ple*.

Even searching within loci of known wound-response genes, we were only ~50% successful at identifying WREs based on conserved clusters AP-1 and GRH consensus sites. Perhaps additional motifs are required within each functional WRE, but they differ between elements, and would thus not be easily identified by comparisons of WREs. The consensus for GRH-like sites may be overly degenerate to accommodate all identified instances within identified WREs, where the binding factor or factors strongly prefer a subset of sites that match the consensus. Spacing between AP-1 and GRH consensus sites does not seem to be of much importance, as they are almost adjacent in *pleSubBMin,* and were brought into close proximity in *DdcΔ123* without severely affecting function, but are quite separate in *msn1.2* and *kkv1* WREs. We note that all identified WREs contain a conserved instance of **G**TGANTCA. This site may assist in identifying further WREs, as well as help unravel *cis-trans* requirement conflicts by indicating which leucine zipper proteins strongly prefer this site, indicating which (presumably bZIP) homo- or heterodimer(s) transduce the wound signal to activate transcription.

The diversity in organization and specific motif sequences resembling AFB and GRH consensus sites in identified WREs is so great that genome-wide searches in FlyEnhancer (Markstein *et al.*, 2002) using consensus motifs and spacing requirements that identify all identified WREs (i.e. 1 GTGANTCA and 1 ACYNGTT, in 450 bp) also identify nearly 4500 clusters that, in general, have no obvious relevance to wound healing. Some fairly stringent searches that exclude one or more identified WREs dramatically reduce the number of clusters, leaving some promising candidates, including the known larval wound responsive gene *puckered*, and genes encoding proteins involved in adherens junctions (*crumbs*), septate junctions (*coracle*), larval cuticle (*Lcp65Ad, stranded at second*), and a second cluster in the 3' end of *kkv*'s first intron. If future genome-scale alignment algorithms improve, identification of true clusters from spurious matches would become much simpler.

Trans-Regulation of WREs

Although we identified FOS and GRH as potential regulators by similarity of required sites to published consensus sequences for these transcription factors, we have no direct *in vivo* evidence that these transcription factors bind the identified sites. Genetic tests of WREs in mutants for these transcription factors only complicated matters, as some WREs are affected, while others apparently function independently of GRH and FOS. In fact, in *grh*^{IM} mutants, *kkv1* is ectopically expressed in the epidermis in late stage embryos. It is unclear whether this indicates that GRH acts as a repressor of *kkv* epidermal expression, or if mutant epidermis/cuticle is weakened to the point that mutant embryos are generating minute tears that only the *kkv1* WRE is sensitive enough to detect.

GRH's optimal *in vitro* consensus binding site is fairly large and specific, but published genomic binding sites often differ significantly from this consensus site. Nevertheless, we have seen that relatively small divergence from the optimal site results in a significant loss of *in vitro* binding. The WREs for *Ddc* and *msn* both contain essentially perfect GRH binding sites, while the *kkv1.2* and *pleSubBMin* GRH-like binding sites match the consensus more weakly. This is consistent with our observations that *Ddc* and *msn* WREs are dramatically affected in *grh*^{IM} homozygotes, while *kkv1* and *pleSubBMin* are not noticeably changed. Perhaps these sites evolved to fine-tune GRH regulation of the wound response, or another transcription factor with similar binding preferences, such as the related CP2 factor encoded by *gemini*, is the *in vivo* regulator of the *ple* and *kkv* wound responses.

Regulation of WREs through sites matching AFB consensus sites is no less complicated. The similarity of required sites to the consensus binding site for AP-1, the bZIP heterodimer of dJUN and dFOS, led us to test the *Ddc -1.4* WRE in mutants for *jun-related antigen* (*jra*), *kayak* (*kay*), and *basket (bsk)* (Mace *et al.*, 2005). We hoped that the new discovery of *sro* as an allele of *kay* (Giesen *et al.*, 2003) would resolve this conflict. Indeed, *pleSubBMin* and *msn1.2* are not activated in *kay*^{sro} mutants, consistent with the poorly differentiated cuticle leading to the *Halloween* class phenotype of *kay*^{sro}. The *kkv1* WRE, however, is not noticeably weakened in *kay*^{sro} mutants.

It is clear, based on all published data, that dJUN/dFOS heterodimers and dFOS/dFOS homodimers would recognize all of these AFB consensus sites *in vitro* (Perkins *et al.*, 1988; Pollock and Treisman, 1990; Zhang *et al.*, 1990). Additionally, other bZIP proteins are known to bind AP-1 consensus sites (Eresh *et al.*, 1997; Masquilier and Sassone-Corsi, 1992), potentially providing *trans*-redundancy that could help explain the complex results in AP-1 component mutants. Future research will attempt to tease out the requirements for different bZIP proteins in regulating these and other wound-responsive genes.

Other than the observations that phospho-Tyrosine and diphospho-ERK are seen rapidly after wounding and that the ERK inhibitor PD98059 reduces *Ddc-1.4* WRE activation, we do not know the upstream signals activating the transcriptional wound response. Both GRH (Liaw *et al.*, 1995; Ylisastigui *et al.*, 2005) and FOS (Ciapponi *et al.*, 2001) are phosphorylated by MAP Kinase, and FOS (along with JUN) is phosphorylated by Jun N-Terminal Kinase (JNK), but it is unclear whether either of these serve to transduce the wound signal or just serve as permissive activators in the epidermis.

Evolutionary Conservation of Wound Response Regulation

The set of identified WREs all share a pair of motifs, matching AP-1 and GRH consensus binding sites. *In vivo* requirements for these sites and their presumed binding factors differ between the elements, but the statistically unlikely event of identifying five separate WREs that all contain conserved sequences matching these motifs strongly suggests some functional relevance. The presence of these sites in these *Drosophila* wound response elements is also interesting because of the widespread requirement of mammalian AP-1 and GRH family members in skin development and wound healing.

Mammalian Grainy head orthologs are expressed in developing skin (Auden 2006), and mutants in *Grhl3* have severe skin barrier defects (Ting *et al.*, 2005; Yu *et al.*, 2006). Additionally, mutants in *Grhl3* are deficient in wound healing, and transglutaminase 1, a key skin cross-linking enzyme, requires *Grhl3* for full expression in the epidermis (Ting *et al.,* 2005; Yu *et al.,* 2006). Microarray analysis of skin in *Grhl3* mutant mice revealed a large set of genes that are altered in mutants, including genes encoding structural proteins of the cornified envelope (the outermost cross-linked skin layer in mammals) and lipid biosynthesis enzymes (Yu *et al.*, 2006).

Similarly, the homologs to *Drosophila* AP-1 factors are required for both skin development and wound healing. Mouse Fos and Jun paralogs are differentially expressed in the different layers of differentiating epidermis (Mehic *et al.*, 2005), and are required in differentiation, proliferation, and migration in various wound healing models (reviewed in (Yates and Rayner, 2002)). The gene *Tgase1*, which encodes the enzyme responsible for crosslinking proteins in the outer cornified envelope, contains sites in its upstream region recognized by GRH (Ting *et al.*, 2005), as well as a site for AP-1 that was shown to be required for full expression in epidermal cells (Jessen *et* *al.*, 2000; Phillips *et al.*, 2004). Mutants in either *grhl3* (Yu *et al.*, 2006) or *c-Jun* (Zenz *et al.*, 2003) have defects in eyelid closure, a process similar to *Drosophila* dorsal closure.

Despite apparent differences in the set of genes activated after wounding in insects vs. mammals, both the overall morphological "mechanism" and the upstream wound regulatory network seems to still be conserved. Identification of additional genes activated after wounding in *Drosophila,* as well as the mechanical and molecular signals that activate this transcription, will likely lead to identification of novel genes or regulatory cascades that are conserved in mammals, potentially aiding in the discovery of novel treatments to aid proper healing.

Acknowledgement

Portions of Chapter III previously appeared in Mace, K.A., Pearson, J.C., and McGinnis, W.J. (2005). An epidermal barrier wound repair pathway in *Drosophila* is mediated by *grainy head. Science* 308, 381-385. I was responsible for the research included in this dissertation.

Portions of Chapter III appear in Pearson, J.C., Juarez, M.T., Lidsky, I., Drivenes, O., and McGinnis, W.J. Common *Cis*-Regulatory Logic of the *Drosophila* Wound Response. *In Preparation.* I am primary author and responsible for the research included in this dissertation.

Materials and Methods

***Drosophila* stocks and genomic DNA:** *D. melanogaster* strain *w*^{1118} was used for germline transformation (Rubin and Spradling, 1982; Spradling and Rubin, 1982), *in situ* hybridization, and a source for genomic DNA. Fly stocks for *D. pseudoobscura, D. virilis, D. immigrans*, and *D. hydei* were supplied by the Tucson *Drosophila* Stock Center (Tucson, Arizona). Genomic DNA was prepared using standard procedures.

PCR: PCR primers generated by IDT (Coraville, Iowa) were used for either classical or inverse PCR, using a standard Touchdown PCR protocol, on genomic or plasmid DNA. Primer sequences available upon request.

Germline Transformation of *Drosophila*: *w*^{1118} embryos were transformed using standard protocols with pH-Stinger expressing DsRed (Barolo *et al.*, 2000; Barolo *et al.*, 2004).

Wounding Procedure: Embryos were collected on apple juice agar plates and aged to 15-17 hrs at 25°C. Embryos were washed into mesh baskets, dechorionated in bleach for 1’, then washed copiously with water. Embryos were then transferred to a clean slab of apple juice agar and aligned for 30-60’ at 18°C, transferred to slides with double-sided tape, then covered in either 1:1 ratio 700:27 weight halocarbon oil. Embryos were then wounded laterally with fresh microinjection needles made from an automated puller, allowed to recover for 3-8 hours at room temperature, and visualized under fluorescent light in either a compound or confocal microscope. Images are representative of at least 2 independent experiments with at least 20 successfully wounded embryos. Pixel intensity levels of images were adjusted for clarity, Adobe Photoshop despeckle, blur, and sharpen functions were used occasionally to enhance clarity. Original images are available on request.

Multiplex Fluorescent *In Situ* Hybridization: Probes were generated from partial or full cDNA clones, obtained from BDGP (Berkeley, CA). Probe labeling and hybridization protocol was as described by Dave Kosman's MFISH protocol (Kosman, 2004).

Sequence Alignments: Sequences cloned by molecular biological techniques or identified Discontiguous MegaBLAST of NCBI Trace archives were trimmed to approximate common boundaries and aligned by T-Coffee (Notredame *et al.*, 2000). Alignments were then adjusted based on evolutionary proximity based on Lalign (http://www.ch.embnet.org/software/LALIGN_form.html) pairwise alignments of small sections of poorly aligned sequences.

Construct boundaries and site alterations:

Ddc-1.4 to-.47: deleted GGCGAGTGGG to GGGAGTCAAG

Ddc-1.4ΔWRE: deleted GGCGAGTGGG to GAGTCCGAGA

DdcΔ1,Δ2,Δ3, and *Δ123* were based on *Ddc-.47.2* (Mace *et al.*, 2005)

DdcΔ1: deleted ACGAGATCGC to ATCAAATTAAG

DdcΔ2: deleted AACTAATTTC to AGTTACTGAT

DdcΔ3: deleted AGCGCCCAAT to GGACTGCGAT

DdcΔ123: deleted ACGAGATCGC to GGACTGCGAT

*Ddc-.47*ETS: changed to

GGATTAATGACG..TCTCTGGCCACA..AGTTGTTAAGCA

*Ddc-.47*TTK: changed to CCGGTAGCTAGGAAT

*Ddc-.47*AFB: changed to CGAGTCCCCCATAA..TTACTCCCCCAGCG

pleSubAMin: AAAGTATCAA to GGAACACGAG

pleSubB: TCTGTGATTG to ATGATTGATGGC

pleSubBMin: TTGGTTTGCA to CGAGGGCTGG

*pleSubBMin*AFB: changed to GTGTGGTGGAGCAC..GCACGGCGCTGACA

*pleSubBMin*CREB: changed to ACGTGGATCAAAAT..GCACGGCGCTGACA

*pleSubBMin*EXD: changed to GCACGGCGCTGACA..AAAATCCCTGCCA

*pleSubBMin*GRH: changed to CACCCGGGAAAGTTG

*pleSubBMin*Hox: changed to

GGAATGGTACTA..CAATACCATACAATGGCCAGCAA..

CTCGTCCGGAACGCACATGGTTGCC..

CTCTTGGTTGTATTTACCGGTTGCGTTTGGTTGAACCATGAATGGTA

TTT

kkv1: CAACAAAGGA to TGGGTGTGTT

kkv2: AAGTGCCAGT to GAGTCCTGTC

kkv1SubA: CAACAAAGGAT to CTCGAAAGAT

kkv1SubB: GCTTACTCCG to ATCAAACCGC

*kkv1*AFB: changed to

GGGTGGTGGATGGC..AAGTGGAGGACTCG..GGAAATCCGCCACAA

*kkv1*GRH: changed to

CAACCTTGGGTCGGC..ATACCTTGGGCTATC..AGACTTTGGGTTTAA.
.CGATCCCCAAGCTTT..TATAGCCAGAGTTG

msn1.2: GAGTGTAGCC to ATTGACAGCA

msn1.3: AGCACTGGCC to GTCTCGTGGA

msn1.2SubA: GAGTGTAGCC to CTCAATTTCC

msn1.2SubB: CCACTGCAAC to ATTGACAGCA

*msn1.2SubB*AP-1: changed to TCCTCTCCCCCACTGG

*msn1.2SubB*GRH: changed to

AATGTCCCAAGGTTG..GAGTTCCAGAGTTC..CAACTGTGGCAAAA

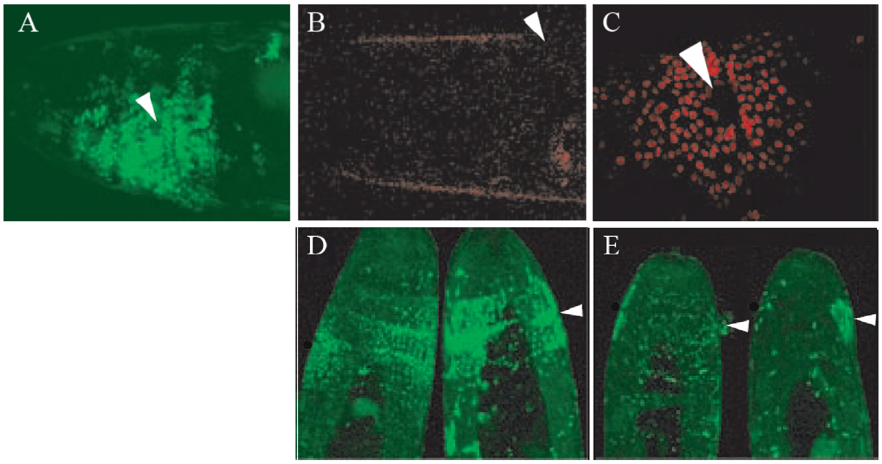


Figure 5. Conserved *cis*-regulatory sequences upstream of *Ddc* require GRH consensus sites and ERK for wound-dependent activation. **(A)** *D. virilis* homolog of *Ddc-.47* (fig. 6c) functions as a WRE in *D. melanogaster*.**(B)** Altering a GRH consensus site in *DmDdc-.47* abolishes WRE activity.**(C)** A 3kb fragment upstream of *ple* with AFB and GRH consensus sites functions as a WRE. **(D)** Injecting DMSO+PBS into the subvitelline space before wounding does not reduce *Ddc-1.4* WRE activity. **(E)** Injecting PD98059, an ERK MAPK inhibitor, dramatically reduces *Ddc-1.4* WRE activity.

Figure 6. Minimal sequence requirements for *Ddc* WRE. (A) Schematic of *D. melanogaster Ddc -1.4* WRE and tested subfragments. Conserved regions with *D. virilis* are indicated by white blocks. Matches to AFB and GRH consensus sites are indicated by **A** and **G**, respectively. Functional WREs are indicated by "+", non-functional elements by "-". All subfragments were tested in DsRed H-Stinger vectors, wounded in parallel to wild-type *Ddc -.47* WRE. **(B,C)** *Ddc-1.4* and *Ddc-.47* both drive GFP reporter expression around aseptic wounds. **(D,E)** Deleting the 46bp CR1 from functional WREs almost completely eliminates activation after wounding. **(F,G,H,I)** Deleting sequences between CR1 and CR2 do not substantially reduce wound-induced reporter expression.

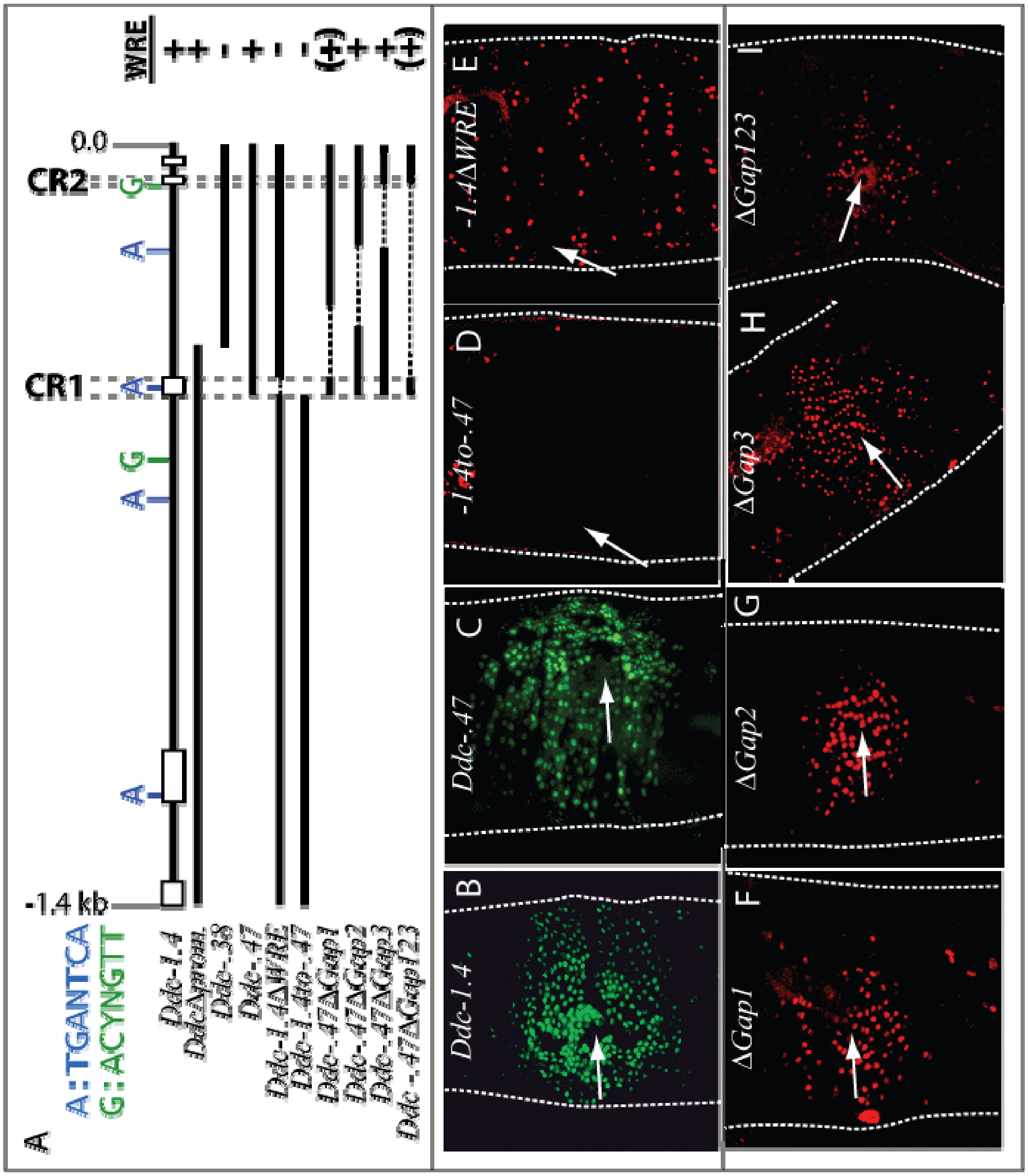

A
A : TGANTCA
G : ACYNGTT
-1.4 kb
0.0
CR1
CR2
A
G
WRE
Ddc-1.4
DdcΔprom.
Ddc-.38
Ddc-.47
Ddc-1.4ΔWRE
Ddc-1.4to-.47
Ddc-.47ΔGap1
Ddc-.47ΔGap2
Ddc-.47ΔGap3
Ddc -.47ΔGap123
+
+
-
+
-
-
(+)
+
+
(+)
B
Ddc-1.4
C
Ddc-.47
D
-1.4to-.47
E
-1.4ΔWRE
F
ΔGap1
G
ΔGap2
H
ΔGap3
I
ΔGap123

Figure 7. Sequences other than AFB and GRH consensus sites contribute to *Ddc* WRE. **(A)** Schematic of *Ddc -.47* WRE and variants with altered binding sites. **(B)** Consensus sequences for known transcription factors matching *Ddc-.47* sites. **(C)** *Ddc-.47* drives reporter expression near epidermal wounds. **(D)** Mutating both AFB consensus sites in *Ddc-.47* abolishes activation at wounds. **(E,F)** Mutating three clustered sequences matching ETS consensus sites reduces, but does not eliminate, *Ddc-.47* WRE activity. **(G,H)** Mutating a conserved sequence matching a TTK consensus binding site reduces, but does not eliminate, *Ddc-.47* WRE activity.

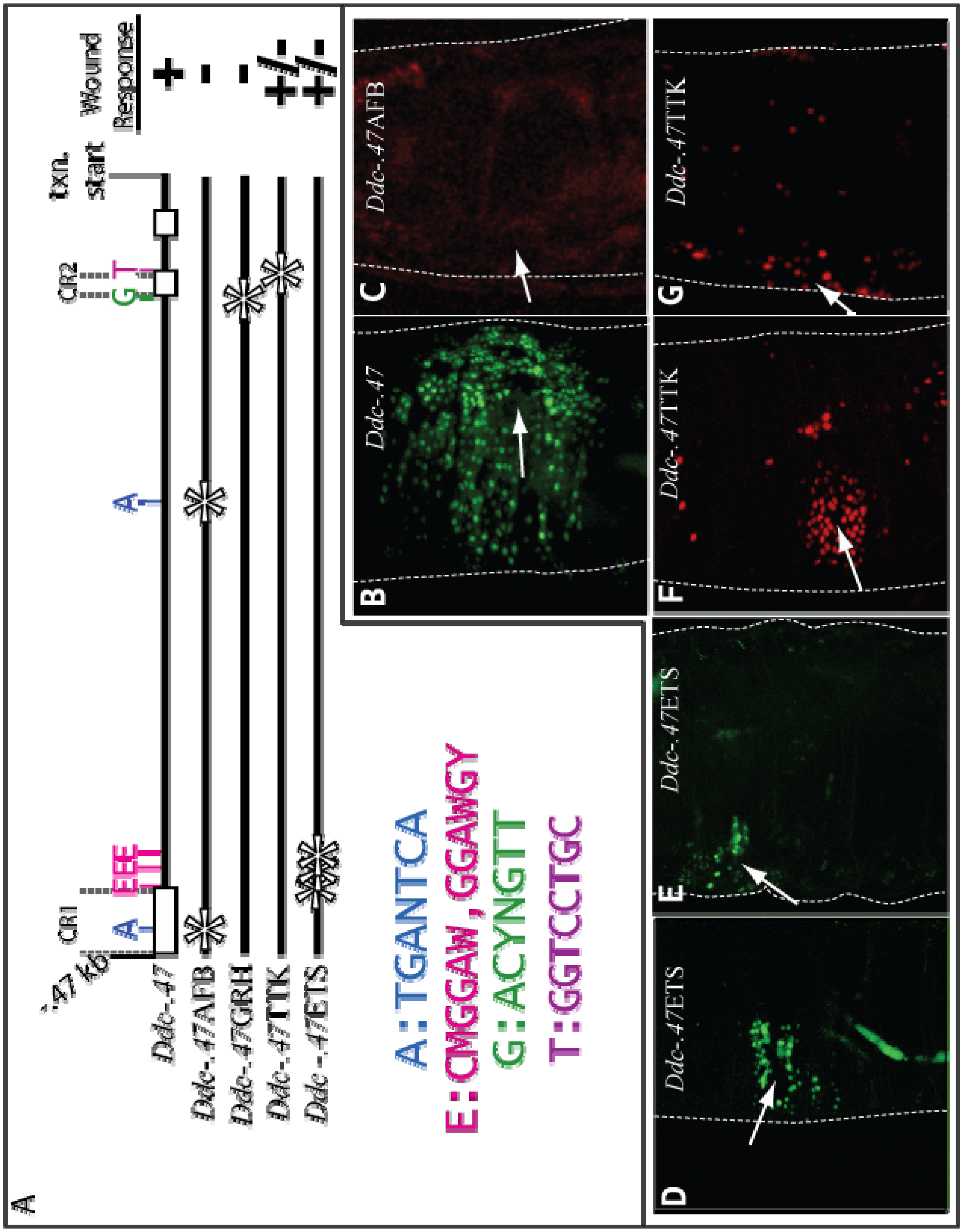
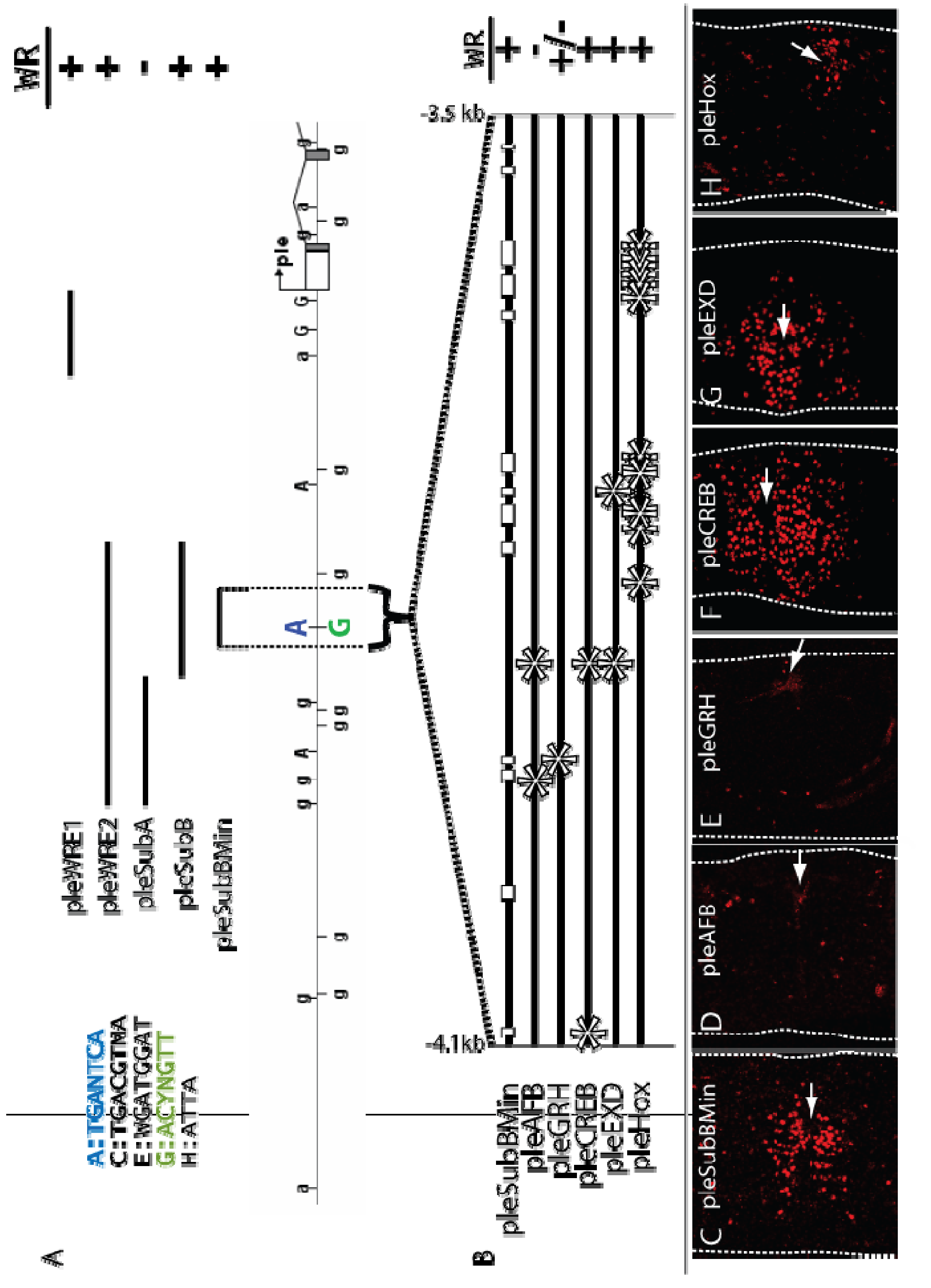
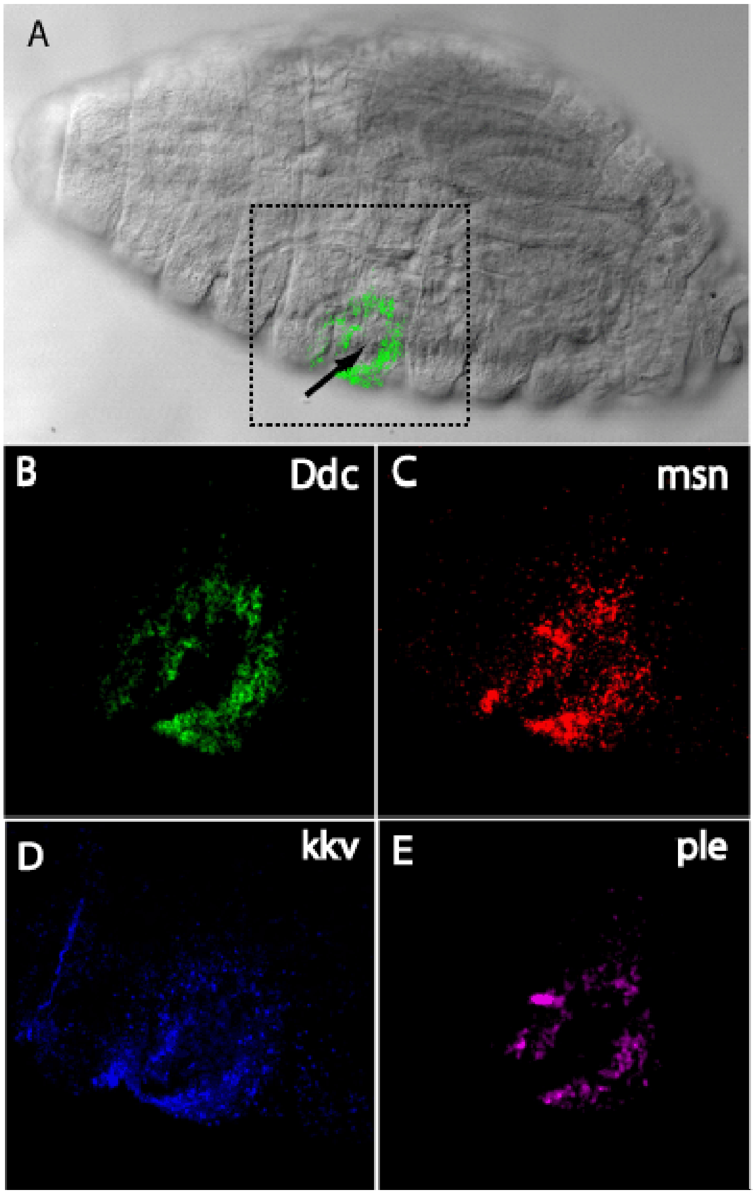

A
-.47 kb
CR1
A
EEE
CR2
G
T
txn. start
Wound Response
Ddc-.47
+
Ddc-.47AFB
-
Ddc-.47GRH
-
Ddc-.47TTK
+/-
Ddc-.47ETS
+/-
A : TGANTCA
E : CMGGAW , GGAWGY
G : ACYNGTT
T : GGTCCTGC
B
Ddc-.47
C
Ddc-.47AFB
D
Ddc-.47ETS
E
Ddc-.47ETS
F
Ddc-.47TTK
G
Ddc-.47TTK

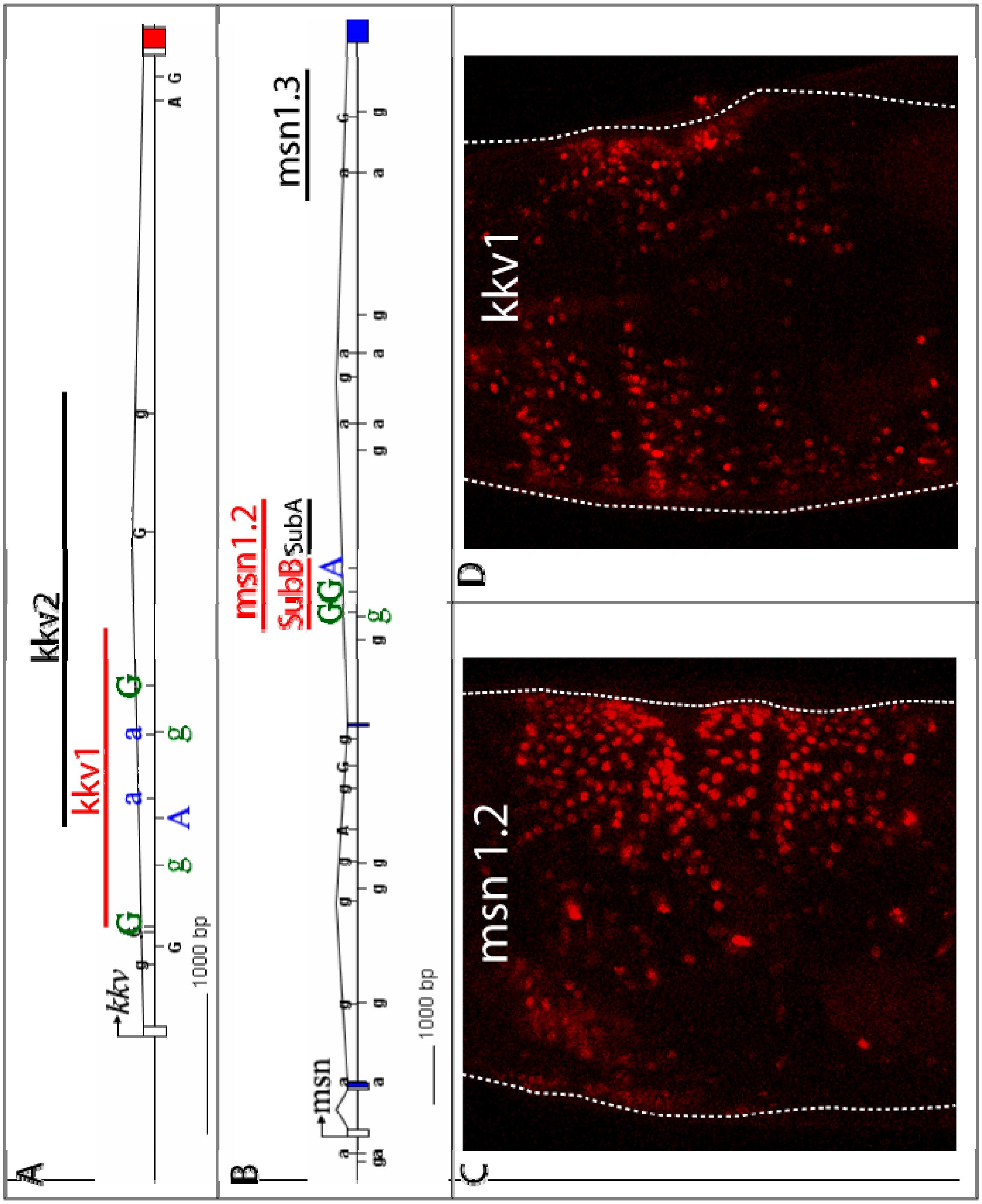
Figure 8. AFB and GRH consensus sites are required for *ple* WRE. **(A)** Schematic of *ple* locus upstream region. AFB and GRH consensus site matches are indicated by **A** and **G**, respectively. Conserved sites are capitalized. A 687bp element containing one AFB and one GRH consensus site functions as a WRE (*pleSubBMin)*. Top left, transcription factor consensus sites identified at conserved positions in *pleSubBMin.* **(B)** Schematic of *pleSubBMin*, with conservation to *D. virilis* indicated by white blocks, and derived elements with mutated binding sites. **(C)** *pleSubBMin* is activated at epidermal wounds. **(D)** Mutating an AFB consensus site and a second AFB/CREB/EXD-like site abolishes *pleSubBMin* WRE activity. **(E)** Mutating a GRH consensus site almost completely eliminates *pleSubBMin* WRE activity. **(F,G,H)** Mutating the AFB/CREB/EXD-like site along with either a CREB-like site or an EXD-like site, or fourteen Hox-like binding sites, has no effect on *pleSubBMin* WRE activity.

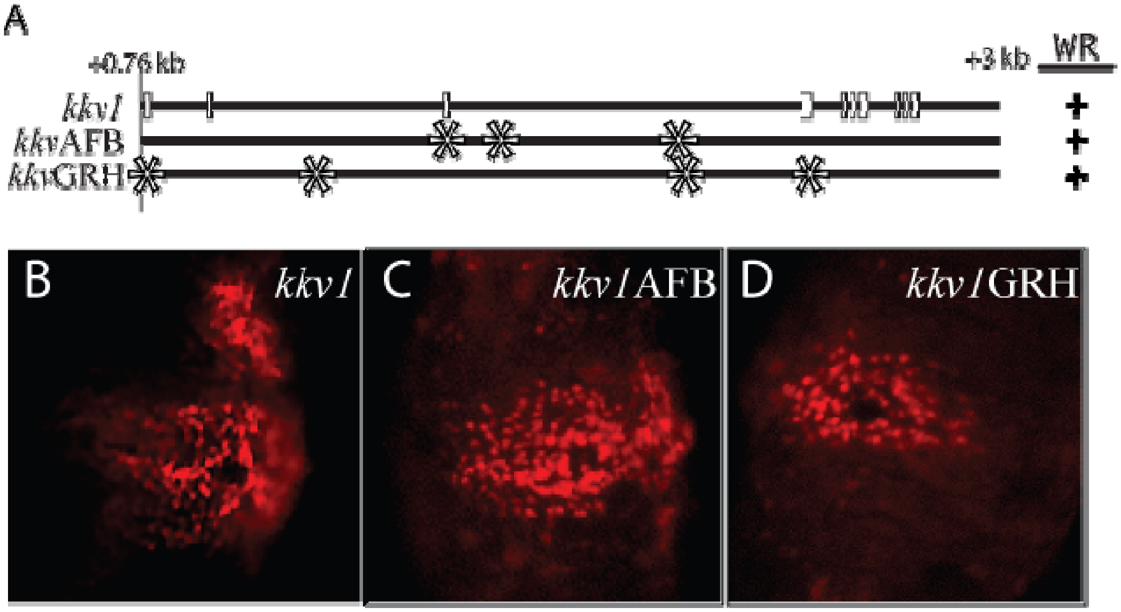

A
A: TGANTCA
C: TGACGTNA
E: WGATGGAT
G: ACYNGTT
H: ATTA
pleWRE1
pleWRE2
pleSubA
pleSubB
pleSubBMin
WR
+
+
-
+
+
ple
A
G
B
-4.1kb
-3.5 kb
pleSubBMin
pleAFB
pleGRH
pleCREB
pleEXD
pleHox
WR
+
-
+/-
+
+
+
C pleSubBMin
D pleAFB
E pleGRH
F pleCREB
G pleEXD
H pleHox



<u>Figure 9. *Ddc, ple, msn,* and *kkv* are rapidly transcribed after wounding.</u> (A) Nomarski image of wounded embryo, fixed 30' post-wounding. *Ddc* wound-responsive expression is superimposed, arrow indicates entry wound, box indicates section imaged in B-E. **(B,C,D,E)** *Ddc*(**B**), *msn*(**C**), *kkv*(**D**), and *ple*(**E**) mRNA were simultaneously detected around an aseptic wound within 30' by labeled antisense mRNA using MFISH (Kosman *et al.,* 2005). No staining was observed around wounds in embryos that were fixed immediately after wounding.

Figure 10. Conserved AFB and GRH consensus site clusters identify *kkv* and *msn* WREs. **(A,B)** Schematics of *kkv* **(A)** and *msn* **(B)** loci, with AFB and GRH consensus site matches indicated by A and G, respectively. Conserved site matches are capitalized. Functional WRE element bounds are indicated by red lines. **(C,D)** *msn1.2* and *kkv1* fragments function as WREs. Both identified WREs contain conserved sequences matching GRH and AFB binding sites.


A
kkv
kkv2
kkv1
G
a
a
G
g
A
g
1000 bp
B
msn
msn1.2
SubB
SubA
GGA
g
msn1.3
1000 bp
C
msn 1.2
D
kkv1



<u>**Figure 11. AFB and GRH consensus site requirements for *kkv1* and *msnSubB* WREs.**</u> **(A)** Schematic of *kkv1* WRE, with conservation to *D. virilis* indicated by white blocks and mutated sites indicated in derived elements. **(B)** *kkv1* wild-type element is activated at wound sites. **(C,D)** *kkv1* is still activated at wounds when AFB or GRH consensus sites are mutated.

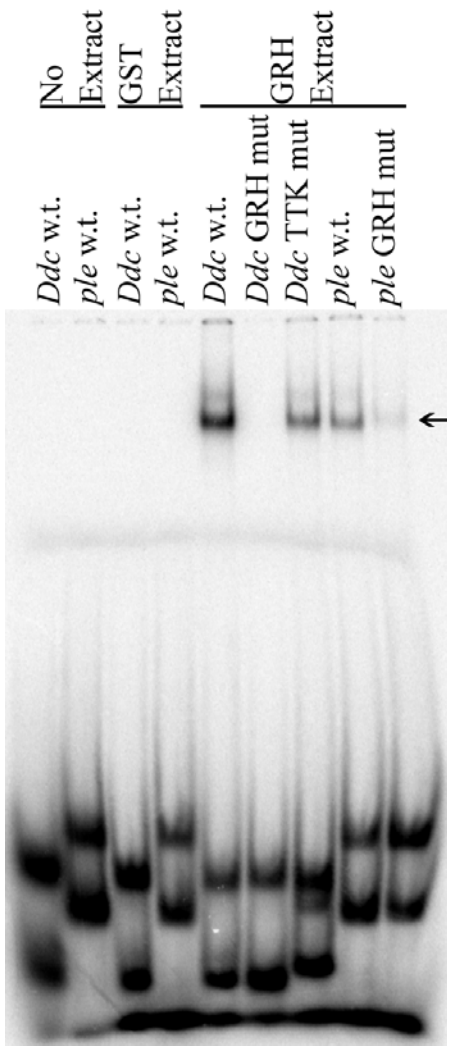


Figure 12. GRH binds required *Ddc* and *ple* GRH consensus sites. Oligonucleotide probes comprising sequences surrounding GRH consensus sites in *Ddc* and *ple* WREs were are bound by *E. coli* crude extract expressing GRH-BE, but not GST. Mutating the GRH binding sites in an identical manner to WRE mutations completely (*Ddc*) or almost completely (*ple*) abolished GRH binding. Mutating the nucleotides within the TTK consensus site adjacent to the GRH site reduced binding affinity by GRH.

Probe sequences:

DdcGRHTTK:	GGGGCGATTG**AACCGGT**CCTGCGGAATTGG
DdcGRHmut:	GGGGCGATT**_CCCAA_GGT**CCTGCGGAATTGG
DdcTTKMut:	GGGGCGATTG**AACCGGT_AGCTA_**GGAATTGG
pleGRH-wt:	GGGGTGATTCAGCACCCA**AACGAGT**TGATCTTGGAAAG
pleGRHmut:	GGGGTGATTCAGCACCC**_GGGAAA_GT**TGATCTTGGAAAG

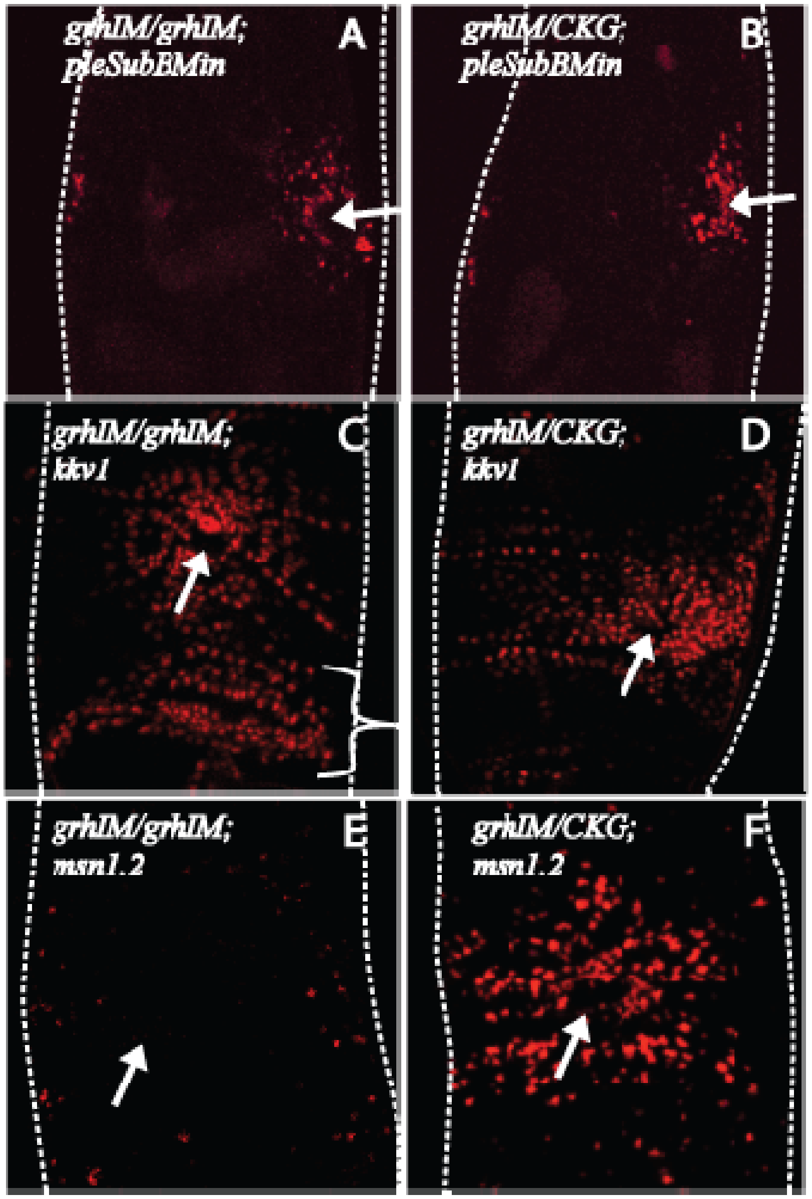


Figure 13. WREs are differentially active in *grh*^{IM} mutants. The strongest lines of *pleSubBMin, kkv1,* and *msn1.2* WREs were introduced into *grh*^{IM} and *kay*^{sro} mutant backgrounds balanced with Kruppel-GFP balancers. GFP^{-} embryos (homozygous mutants) were compared to GFP^{+} embryos (heterozygotes and homozygotes for balancer) for WRE induction. **(A, B)** No significant difference was observed in extent of *pleSubBMin* activation in GFP^{-} compared to GFP^{+} embryos. **(C,D)** *kkv1* WRE is ectopically activated in dorsal/lateral epidermis in unwounded *grh*^{IM} homozygous embryos (data not shown), but no change in wound-induced activation was observed. **(E,F)** *msn1.2* wound-dependent activation is dramatically reduced in *grh*^{IM} mutants.

Chapter IV:

Conclusions/Final Thoughts

While the set of transcription factors binding DNA regions is the likely most fundamental mechanism for gene regulation, the extent of knowledge as to how this works seems to consist primarily of a growing collection of disparate examples of how different genes are regulated *in cis.* Some patterns do emerge, at least in the minds of scientists who study *cis*-regulatory elements. Following is a selection of these assumed biases that I applied in my research, followed by contributions (if any) of my results.

1. Functional *cis*-regulatory elements are independent and separable. The group of binding sites that work together, when bound by appropriate transcription factors, to cause expression of a gene in one tissue, is clustered in one small area of the genome (~500 bp, another assumption) outside of other groups of binding sites that control other aspects of the same gene's expression.

I found that, in most cases, sub-elements chosen based on clusters of important motifs would maintain expression of interest, while progressively losing other expression as the elements got smaller. In some cases, no discernable difference in expression was noticed between larger and smaller elements (*e.g. Dll304Ex* vs. *Dll304Min*), while other elements were significantly "cleaner" as I generated smaller sub-elements (*pleWRE2* vs. *pleSubBMin*, *msn1.2* vs. *msn1.2SubB*). I never identified an element of *ple* that separated anal pad expression from wound expression, but results from site mutations suggest that this could be done.

The notable exception is *kkv1*, where neither tested subelement, both of which contained the assumed-to-be-required AFB and GRH consensus sites, maintained any wound response activity. In retrospect, this is not surprising, as the AFB and GRH sites are not required in the *kkv1* WRE, so I may have just tested the wrong subfragments. Alternatively, a small region of *kkv1* that was not incorporated into a tested subfragment could contain WRE function.

2. *Cis*-regulatory elements are better conserved than surrounding sequences. *Cis*-regulatory elements are comprised of a set of binding sites, and these binding sites are required for transcription factors to exert their positive and negative influences on transcription. Assuming no changes in transcription factor expression or binding affinity, to maintain target gene expression, it is assumed that either the original binding site must remain or a compensatory site must evolve nearby to provide redundancy. Since most binding site sequences are complex enough that they cannot easily spontaneously come into existence by random mutations, evolutionary pressure keeps the original binding site, therefore it is conserved, and thus the set of binding sites is conserved along DNA, and the *cis*-regulatory element as a module is more conserved than surrounding non-functional sequence.

This assumption turned out to be consistently and remarkably true in my research. Some regulatory elements were significantly more conserved as a whole than others (compare *Dll304Min* to *Ddc-.47*), but functional elements were always contained within “islands” of conservation, and the required motifs were conserved, while tested non-conserved sequences had minimal effect on expression (*Ddc-.47* vs. deletions).

This assumption was especially effective in identifying two independent, novel motifs required for *Dll* embryonic expression. Even though they are not the most highly-conserved motifs in the region, the combination of assumptions (1) and (2) led to the filtering of both non-conserved (non-essential?) sequences and motifs that are likely important for other limb-independent *Dll* expression.

3. Co-expressed genes can share regulatory mechanisms. This assumption underlies any algorithm that compares sets of *cis*-regulatory elements for common motifs, and to a degree any algorithm that searches for clusters of a given motif on a genome-wide scale, and indeed even the use of phylogenetic conservation of sequences (*i.e.* alignable sequences are there because the expression pattern has been maintained in related species, due to common regulation).

Searching for clusters of AFB and GRH consensus sites near other wound-response genes identified two *ple* WREs, as well as WREs for *kkv* and *msn*. Granted, the *kkv* WRE may not actually *need* those sites, but this assumption has worked fairly well. Other examples abound in the literature, but this may be due to a bias in the genomics age of being able to find examples confirming the validity of this assumption because this assumption makes it so much easier to find fitting *cis*-regulatory elements. It is much more complicated to prove that co-expressed genes are *not* co-regulated, although the *kkv1* WRE may be on the way to doing so.

TWINE (Appendix A) was written to take advantage of assumptions (2) and (3), separately or in combination. It was not a particularly sophisticated implementation of this, but does succeed at identifying a certain number of motifs above background in both the *Dll* limb and WRE paradigms, due largely to the significantly reduced search space because of phylogenetic footprinting and the tendency for functional motifs to be repeated both within and between functional elements with similar expression.

I strongly suspect that, despite certain "complications" with *cis* vs. *trans* regulation, I have been lucky in choosing paradigms for studying *cis*-regulation, having identified multiple *cis*-regulatory elements and a large subset of their required components. While consistency makes for a better story, the complications, and the resolution of those complications, provide years/decades of interesting research.

As for identifying any common rules of co-regulated genes, large sample sizes, evolutionary conservation, and ease of identification of specific cell types is a major advantage. One system that seems to have both of these is ventral midline neural and glial cells in insects. Recent publications have cataloged the expression patterns of hundreds of genes expressed in subsets of these cells. The identification of *cis*-regulatory elements controlling these expression patterns may reveal extensive shared logic, or may demonstrate that multiple independent regulatory pathways can be used to achieve identical expression of large sets of genes.

As genome-scale sequencing becomes progressively cheaper and more groups get their “pet species” sequenced, more complete pictures will emerge of the evolutionary history of *cis*-regulatory elements, which will then lead to more efficient methods for analyzing their functions. More cataloging of exceptions, as well as adherents, to established assumptions, will hopefully finally establish the fundamental importance of “junk DNA”, and allow people to feel more at ease with their inadequate gene counts.

Appendix A

TWINE:

A Java Program for Simple Graphically-Assisted Searches of Repeated and Conserved *Cis*-Regulatory Motifs

Computer-aided searches for over-represented sequence motifs have become an essential component of *cis*-regulatory analysis (GuhaThakurta, 2006). Since functional regulatory elements often contain multiple instances of required transcription factor binding sites, simple searches for clusters of motifs that are statistically unlikely to occur in random sequence can indicate functional relevance. Conservation of sequences between homologous regulatory elements also often indicates important motifs, as functionally relevant DNA sequences are less likely to change during evolution compared to neutral sequence (Mirny and Gelfand, 2002). *Cis*-regulatory elements for different genes with similar expression patterns sometimes share regulatory logic, so comparisons of functionally similar elements can reveal common motifs that bind to common transcription factors (Erives and Levine, 2004; Markstein et al., 2004; Senger et al., 2004). To take advantage of all of these *cis*-regulatory "rules", I wrote TWINE, a Java program that utilizes evolutionary conservation to perform comprehensive searches for over-represented motifs in one or more *cis*-regulatory elements with shared expression patterns.

Input

TWINE input is a FastA-formatted sequence of a known *cis*-regulatory element (Reference Sequence), aligned to identifiable homologs. Each opened alignment is an Aligned Sequence Object (ASO). Since linear conservation contributes greatly to the function of TWINE, manually edited/optimized multiple alignments are recommended. Multiple aligned regulatory elements can be opened, and all opened ASOs will be used in motif searches.

Pseudo-multiple alignments can be generated from multiple pair-wise alignments between a common reference sequence and its various homologs. For example, given a set of alignments of *D. melanogaster* sequence to *D. pseudoobscura, D. virilis,* and *D. mojavensis* downloaded from VISTA web genome browser (http://pipeline.lbl.gov), a FastA file containing these alignments in the order *Dmel Dpse Dmel Dvir Dmel Dmoj* (Reference Sequence alternating with homologs) can be aligned using "Make Vista Multi-align", outputting a file containing the reference sequence *Dmel* aligned to all of its homologs in the order *Dmel Dpse Dvir Dmoj.* This alignment can then be optimized by manual alignment and used in TWINE searches.

Initial Search

TWINE performs a comprehensive search of all motifs of a user-defined length contained within the Reference Sequences of all opened alignments, noting all matches to each motif within these opened sequences and aligned orthologs. If a motif finds matches in enough sequences and these matches are conserved in enough homologs (linearly or non-linearly), this motif is saved.

Search Settings

After all desired alignments are opened, but before searching for motifs, settings can be adjusted to suit user needs.

- Window size: Motif size to search. For each overlapping window along each open Reference Sequence, a motif is generated for searching in all open sequences.

- Max. nucl. Conc: The maximum percentage of any nucleotide A,C,G,T, allowed in a motif to be considered. A primitive simple-sequence filter.
- Max. mismatches: The maximum number of non-matching nucleotides between the currently tested motif and any potential match within open sequences.
- Max. Match uncons. Nucs: For a given match to the currently tested motif, the maximum number of nucleotides that is not linearly conserved in the alignment for the match to be considered linearly conserved.
- Min. Match Cons. Level: The minimum percentage of nucleotides at each match alignment position that is linearly conserved for the nucleotide position to be considered linearly conserved.
- Min. Matches Per ASO: The minimum number of matches to the currently tested motif within an Aligned Sequence Object (ASO) for the ASO to be considered as containing matches. Useful for requiring 2 or matches of a motif within each ASO.
- Min. Num. ASOs with Match: The minimum number of ASOs (among opened ASOs) required to have matches for the motif to be saved.
- Min. Num. Cons. ASOs with Match: The minimum number of ASOs required to have conserved matches (based on above settings) for the motif to be saved.
- Min. P-value: Maximum P-value, as calculated based on Poisson distribution, of the set of identified motif matches for the motif to be saved.

Once settings are adjusted, clicking “Analyze!” will search all ASOs for the set of motifs that meet user settings. Any motif that meets these requirements will be displayed in the Motifs box.

Additional restrictions can be placed on the set of motifs generated for searches, utilizing linear conservation to both reduce the motif set and augment the motif sequence used in the search. Selecting Options>Change Parameters>Change reqs. for generated motifs will bring up a window called “Set Motif requirements”. Selecting the “Use these parameters in searches” checkbox will activate this function. When activated, only motifs within a window that meets or exceeds user-set requirements for minimum conservation within the window will be used for searches. Additionally, any nucleotides within the motif window that are not perfectly conserved will be converted to degenerate nucleotides for the search. When matches are scored relative to this degenerate motif, those matches that only differ from the Reference Sequence motif source at highly diverged nucleotides will get a higher match compared to matches that differ at conserved positions.

Viewing Motifs

Selecting a motif in the Motifs box displays all matches to the reference sequence in the currently selected ASO. Selecting one of these matches will display the alignment of sequences to this match in the selected ASO. A separate box displays all matches to the selected motif in all species and all ASOs. A frequency matrix of all matches to the selected motif is displayed for reference, along with a consensus sequence using IUPAC degenerate nucleotides as needed.

The top box displays a conservation plot of the currently selected ASO, ranging from 0 to 100% conservation at each nucleotide of the alignment, each pixel representing one nucleotide of reference sequence by default. The number of nucleotides per pixel can be adjusted by the Zoom slider. The number of consecutive nucleotides calculated in conservation levels at each pixel can be adjusted by the Blur slider. Matches to the currently selected motif in the displayed ASO are indicated by color-coded vertical lines.

Auto-generated motifs may be sorted by selecting the desired sort method, then selecting Update. Motifs and matches may also be filtered to exclude or include motifs or matches that are deemed “conserved” based on user specifications. The definition of “conserved” may be changed from including non-linear conservation to only considering linear conservation.

Custom motifs can be entered using either strict or degenerate consensus sequences, and will be saved in a separate box. The same options for auto-generated motifs are available for custom motifs, but match restrictions are increased, while global restrictions are reduced, by default.

Optimizing Motifs

Custom and default motifs can be optimized automatically to find the best degenerate motif that meets user specifications. For the selected motif, all possible degenerate motifs will be generated and searched, substituting a user-defined number of N's (x) at all possible positions of the motif, where x is "Max. degeneracy" in the "Optimizing parameters" menu, *e.g.* CAATTAA, x=3 generates NNNTTAA, NNANTAA, NNAATAA, *etc*.

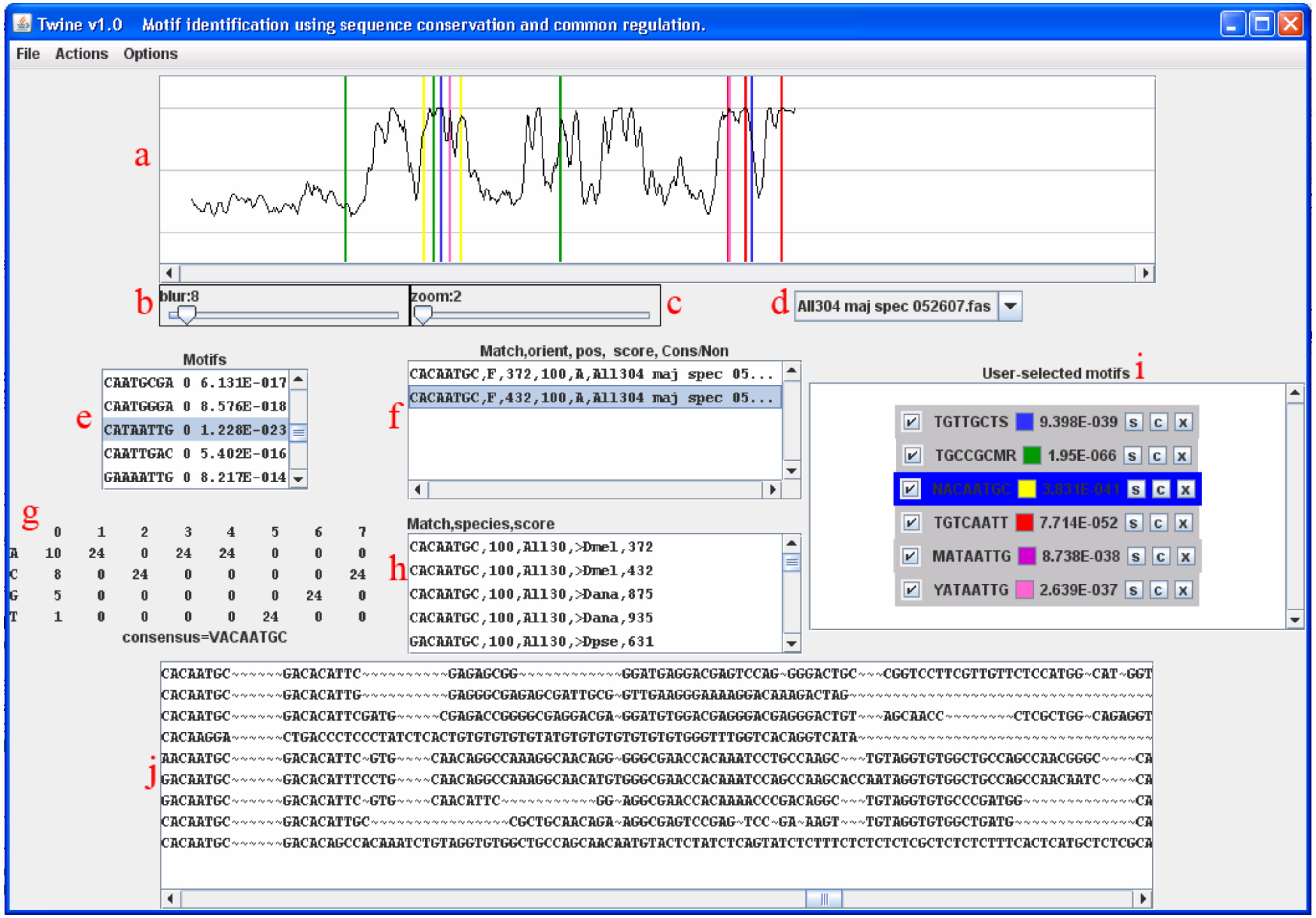
Once the search is complete using all generated degenerate motifs, all maximally degenerate motifs that find sufficient matches to pass requirements set in "Optimizing parameters" are displayed in a pop-up window. The user may select one or more of these motifs to "regenerate", *i.e.* find the least degenerate motif that still meets "Optimizing parameters". Optimized motifs are displayed in order of increasing degeneracy, and the user may select one or more optimized motif to be added to the Custom Motif box by holding down Ctrl (PC) or Option (Mac).

Random Final Notes on TWINE

In addition to serving as a simple conservation-based motif search program to identify novel motifs in multiple co-expressed and putatively co-regulated sequences, TWINE can serve as a method to visualize the organization of motifs relative to sequence conservation. Thus, given a set of motifs presumed to be required in a putative *cis*-regulatory region, TWINE can be used as a visualization tool for intelligent determination of likely boundaries of *cis*-regulatory elements based on regions of conservation and clustered motifs.

The current statistical and searching algorithms are simple, intuitively obvious automations of sequence analysis techniques. A cornucopia of more accurate, robust, and quicker algorithms exist and should be implemented ((Jones and Pevzner, 2004; Siddharthan et al., 2005; Sinha et al., 2004) come to mind). Additionally, customizable importing/exporting of motif and match objects would enable users to analyze data generated from other searches in TWINE, or visualize data from TWINE in other visualization programs such as GenePalette (Rebeiz and Posakony, 2004).

Figure 14. TWINE main window. A screenshot of TWINE in Windows XP, after an analysis of an alignment of *Dll304* and 3' conserved regions, with several automatically optimized motifs from the search.
a. Conservation plot displays percent conservation at each position of the Reference Sequence to aligned homologous nucleotides of the currently selected Aligned Sequence Object (ASO).
b. Blur slider controls the number of nucleotides of Reference Sequence conservation to average at each point on the x-axis.
c. Zoom slider controls the resolution of the conservation plot.
d. Dropdown menu allows selection of which open ASO to display in conservation plot box and other ASO-specific boxes such as match (**f**) and alignment (**j**) boxes.
e. Motifs box displays all motifs found in the current search, sorted alphabetically by forward and reverse sequence.
f. Match box displays all matches to the selected motif in (**e**) or (**i**) in the currently selected ASO. Orientation (F/R), position in Reference Sequence, percent match to motif, Type of conservation (C=linear, A=non-linear, N=not conserved), and sequence in which the match is found.
g. Position Weight Matrix of all matches to the selected motif in (**e**) or (**i**) in all species, and a consensus sequence derived from the matrix.
h. All Species Match box displays all matches to the selected motif in all species, in all ASOs.
i. User-defined motifs contain optimized motifs and manually-inputted motifs. Multiple colors can be selected to differentiate motifs, and individual motifs can be hidden from display in (**a**) by using the checkbox. Pressing the "**s**", "**c**", or "**x**" will allow motif settings to be changed, the motif to be copied, or deleted, respectively.
j. Alignment of currently selected ASO, adjusted so that the left-most part of the viewed alignment is the selected motif from (**e**) or (**i**), or any position selected from the conservation plot in (**a**).


Twine v1.0 Motif identification using sequence conservation and common regulation.
File
Actions
Options
a
b
c
d
e
f
g
h
i
j
S
C
X

blur:8

zoom:2

All304 maj spec 052607.fas

Motifs

```
CAATGCGA  0  6.131E-017
CAATGGGA  0  8.576E-018
CATAATTG  0  1.228E-023
CAATTGAC  0  5.402E-016
GAAAATTG  0  8.217E-014
```

Match,orient, pos, score, Cons/Non

```
CACAATGC,F,372,100,A,All304 maj spec 05...
CACAATGC,F,432,100,A,All304 maj spec 05...
```

User-selected motifs

☑	TGTTGCTS	🟦	9.398E-039
☑	TGCCGCMR	🟩	1.95E-066
☑	NACAATGC	🟨	3.831E-041
☑	TGTCAATT	🟥	7.714E-052
☑	MATAATTG	🟪	8.738E-038
☑	YATAATTG	🟪	2.639E-037

```
    0    1    2    3    4    5    6    7
A  10   24    0   24   24    0    0    0
C   8    0   24    0    0    0    0   24
G   5    0    0    0    0    0   24    0
T   1    0    0    0    0   24    0    0
consensus=VACAATGC
```

Match,species,score

```
CACAATGC,100,All30,>Dmel,372
CACAATGC,100,All30,>Dmel,432
CACAATGC,100,All30,>Dana,875
CACAATGC,100,All30,>Dana,935
GACAATGC,100,All30,>Dpse,631
```

```
CACAATGC~~~~~~GACACATTC~~~~~~~~~~GAGAGCGG~~~~~~~~~~~~GGATGAGGACGAGTCCAG~GGGACTGC~~~CGGTCCTTCGTTGTTCTCCATGG~CAT~GGT
CACAATGC~~~~~~GACACATTG~~~~~~~~~~GAGGGCGAGAGCGATTGCG~GTTGAAGGGAAAAGGACAAAGACTAG~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
CACAATGC~~~~~~GACACATTCGATG~~~~~CGAGACCGGGGCGAGGACGA~GGATGTGGACGAGGGACGAGGGACTGT~~~AGCAACC~~~~~~~~CTCGCTGG~CAGAGGT
CACAAGGA~~~~~~CTGACCCTCCCTATCTCACTGTGTGTGTGTATGTGTGTGTGTGTGGGTTTGGTCACAGGTCATA~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
AACAATGC~~~~~~GACACATTC~GTG~~~~CAACAGGCCAAAGGCAACAGG~GGGCGAACCACAAATCCTGCCAAGC~~~TGTAGGTGTGGCTGCCAGCCAACGGGC~~~CA
GACAATGC~~~~~~GACACATTTCCTG~~~~CAACAGGCCAAAGGCAACATGTGGGCGAACCACAAATCCAGCCAAGCACCAATAGGTGTGGCTGCCAGCCAACAATC~~~CA
GACAATGC~~~~~~GACACATTC~GTG~~~~CAACATTC~~~~~~~~~~~GG~AGGCGAACCACAAAACCCGACAGGC~~~TGTAGGTGTGCCCGATGG~~~~~~~~~~~~~CA
CACAATGC~~~~~~GACACATTGC~~~~~~~~~~~~~~~~CGCTGCAACAGA~AGGCGAGTCCGAG~TCC~GA~AAGT~~~TGTAGGTGTGGCTGATG~~~~~~~~~~~~~~CA
CACAATGC~~~~~~GACACAGCCACAAATCTGTAGGTGTGGCTGCCAGCAACAATGTACTCTATCTCAGTATCTCTTTCTCTCTCGCTCTCTCTTTCACTCATGCTCTCGCA
```

Appendix B

Alignments and Annotations of

Cis-regulatory elements and *Drosophila* orthologs

Dll304 and adjacent conserved sequences

```
           |--Dll304 Start-->
Dmel  1    GAATTCCCA~AACT~GGTGGAG~~~~~TG~GCTATCGG~ATCGGTCTGTCAAAATGG~TG
Dsim  1    GAATTCCCA~AACT~GGTGGAG~~~~~TGGGCTATCGG~ATCGGTCTGTGAAAG~GG~TG
Dyak  1    GAATGCCCA~AAGT~GCTGCATGGATGTG~GCTATCGG~TTTTTG~~~~~~AAAGGG~TG
Dana  1    ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dpse  1    GAATACCCT~ACGC~AGGCGGG~~~~~GA~GTTGGGGA~GCATCATCTGGGACTGAC~TG
Dper  1    GAATACCCTTACGCCAGGCCGG~~~AGGAGGTTGCGGAAGCATCATCTGGGACTGACCTG
Dwil  1    ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dhyd  1    ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dmoj  1    ~~~~~~~~~~~~~~~~~~~~~~~~~~~TCTATTCCAGT~TTTGTTGAAAGTTTTTTT~TT
Dvir  1    ~~~~~~~AAGCTTATTTTAGGAATGTAAT~TGCTTGGA~TTAAGCGCAAGTTTAGTT~GG
Dimm  1    GGGCCACTTTTGCCGCCACGCAAACACGCCATGGAACG~AGTCTGTTAGATTTTTGT~TT
Dgri  1    ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Sleb  1    ~~~~~~ATCTCAATTAATATNTACTATAAGTTAAACAA~TTCACTCCCAGAGGGGTC~AT

Dmel  51   TA~TTTGCA~GGTACAGTGTTTCATTTCCGCAC~~AAAAACTGAGTTTG~~~~~~ATAAG
Dsim  51   TA~TCTGCA~GGTACAGTGTTTCATTTCCGCACAGAAAATGCGAGTTGG~~~~~~ATAAA
Dyak  50   TA~CCTGCT~GGTACAGTGTTCCATTTCCGCACAA~~~~~~~~~~~~~~~~~~~~~~~~~
Dana  1    ~~~~~~~~~~~~~~~~~~~TCCAATTTCCGATCCATAT~~~~~~~~~~~~~~~~~~~~~~
Dpse  51   GC~TGGGAC~ACTTGGGCCGAATGGAAAGGTTG~~TAAAA~~~~~~GTAGGTGAGTGGGA
Dper  58   GCCTGGGACCACCTGGGCCGAATGGAAAGTATGGATCATCGTCACTGTAGGTGGATTGGG
Dwil  1    ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dhyd  1    ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dmoj  32   TT~AAAAAT~TAAGAACATTTTTAGATTATTTATATGTAAATATTTTAA~~~~~~TGAAG
Dvir  50   ~T~CTATTT~ATTTTTCTATTTATATTATGCTA~~ATCGAAATTGTCTT~~~~~~AAACT
Dimm  59   GG~CTCTGC~GACTATTTCTATTCGTAAACTGGTCTGAGTATTAGACAG~~~~~~ATTCA
Dgri  1    ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Sleb  53   CG~AGCTGA~ATTCGTAGGTCTCTTGAATTCGTACGTGATTTGTGAAAT~~~~~~TGTAA

Dmel  101  TAAGGCTAGTTTTACTAATTTCTTCAAACCG~~TCTA~~~~~~~~~~~TAACATCCACAC
Dsim  103  TAAGGCTAGTCTTACTAATTTGTACATATGTGATCTA~~~~~~~~~~~TACCATCCACAC
Dyak  82   ~AAAGCGAGTTTTACTAATTTCTGTATATCCAGAATATATTAATACCCTACCAGCTACAC
Dana  19   ~~~~~~~~~TTTTTCAAATTTTTGTGGGAAAATCAAAGATTCTCATTTGTGCCCCTTGGG
Dpse  101  A~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dper  118  A~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dwil  1    ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dhyd  1    ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dmoj  84   TTTAAACTGTAAATAAGAAAATGTAATATTTTTAATCCAATTTAGAAAAAGAAATCAATG
Dvir  100  AAGATCGAAAATCTCTTAAAATTAGACGAAA~~AGTC~~~~~~~~~TTACGTTAAGAACT
Dimm  111  TAGTCGTATTCGTAGTCGTATTCGTTTTCGT~~ATTC~~~~~~~~~~~GTTTCCGTTTCA
Dgri  1    ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Sleb  105  AAGATTGCCAATTCCTAAACGAAAGTCCTACTGCAAATAAGGCTTTTAAAAATTAACAGA
```

```
Dmel  148  CGAATTTCGCCTTATGGCTTAAGGTCGTCGAAGGTGCTCGAAATACCCGCAAATGGACAT
Dsim  152  CGAATATCGCCTTATGGCTGT~~~~~~~~~AAAGTGCTCAAAATACCCGCAAATGGACAT
Dyak  142  CGAATATCGCCTTATGGCTTAAGGTCGAGGAAGGTGCTCGAAATACCCGTAAGTGGGCAT
Dana  71   GTAGTTATATCTCGGCCAATTCCTGCCCGATTCTTGGGCGGAATACCC~TAAAC~GGTTT
Dpse  101  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~CGAAAAATGACAAAATGGAACT
Dper  118  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ACAAAAAATGCCAAAATGGAACT
Dwil  1    ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~AAGGTGGTTAATATACCAGTTAATAGATTT
Dhyd  1    ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dmoj  144  AAATCTAAAACAAAAAACTTGAGTAAAATTAACTGCAAAATTCTAGTAAATTCAAAAAGT
Dvir  149  TATTTGGAAAACTGATTTAATACAGGAAAATATATTTTGAGTTTGACTTTTCCTTGAATG
Dimm  158  GAGACGCCAACATGGCGACACA~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dgri  1    ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Sleb  165  AAATGAAAAAGATGGTCTTTAAGAGTGTCCAAACATATTATCAACAGAGTGTGACCTTTT

Dmel  208  GTGGAGAGAGGAGC~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dsim  203  GTGGAGAGAGGAGC~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dyak  202  GTGGAGAGAGGAGC~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dana  129  GTGGATCGATTCCTCATCGATTTCCATCAAAACCTCAACACAAATTT~~~~~~~~~~~~~
Dpse  124  GAGGATATATTAGAGCCCAAACTACAGTTTATTAAAGGTCT~~~~~~~~~~~~~~~~GAC
Dper  142  GAGGATATCTACAGAAGGTCTAACCAATATCTCCACGACTGAAAACGTCCCAGCAGA~~~
Dwil  31   AAATATAAAAAGAATTTCTTCATGTATGATAATCCCAACAAGAGATAATCTGCAATTGCC
Dhyd  1    ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dmoj  204  CTTTGCTTTACATAATATAAATATACAATAAATTTTCTCTAAGTGTA~~~~~~~~~~~~~
Dvir  209  CTTATTTTAAAAAATCCTTGTCGGGCGTCTCTTTTTCTCGCTGTGTA~~~~~~~~~~~~~
Dimm  179  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dgri  1    ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Sleb  225  AAACTTACACCCCTTTTTATTGAAGTGTACGCCAGGTACAGCAACTA~~~~~~~~~~~~~

Dmel  221  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dsim  216  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dyak  215  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dana  175  ~~~TTGGGACGAAATTTGAAGAGGATATTGAGAATATTGAGAGGATACCTTAAACGCTTT
Dpse  168  CAATGATATATCTACAGAAGATATATCAAATGATATACATAGACGCATCTAAAAGGGTCT
Dper  198  ~~~~~~~~~~~~~~~~~~~~~~~~ATCAAATGATATACATAGACGCATCTAAAAGGGTCT
Dwil  91   TATTCATGAATAACAACTTATGAATATATGTATTCAGATTCACATTAAAGAAAGACTATT
Dhyd  1    ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dmoj  250  ~~~GCTACGCTTTTGCGTGTTTTTAGTCAGAGAACTT~~~GGGCCG~~CAGTTG~~~GTG
Dvir  255  ~~~GCTTTGGTTTTGCGTG~~TTTGGTCAGAGAACTTCTTGGGCCA~~CACTTG~~~GTG
Dimm  179  ~~~~~~~~~~~~~TTTGGCTTGCGTGCCTTAGAACTT~~~GGCCAATGCTCTTGTCTGTG
Dgri  1    ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Sleb  271  ~~~AGCGCGGAAAGTGAATTGAGTGAGAGGCTGGGCTAAAAGTGTATCAAAGCCATTTTT

Dmel  221  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dsim  216  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dyak  215  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dana  233  CTAGATCGAGCCCCCATCGATCTGCCCCATTGAACTTTGACACAATTTTCGCCAGAAAAC
Dpse  228  AAGTGTAAGAGAAACACAACGTTAAAAAGATCAGAATGTAAACAGATAAAAGATAACCGT
Dper  235  AAGTGTAAGAGAAACACAAAGTTAAACAGATCAGAATGTTATCAGATAAAAGATAACCGT
Dwil  151  GAAAAATCATAAATCTGTTGATTTGGATTTTTTAGCAAGTTTTTTTCTCTCTTTATGTGG
Dhyd  1    ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dmoj  300  GCTTT~~GGCTATGACATAAGTGC~~~~~TTAGAGC~~~~~~~~~~~~~~~~~~~~~~~~
Dvir  306  GCTG~~~GGCTATGACATAAGTGC~~~~~TTAGAGC~~~~~~~~~~~~~~~~~~~~~~~~
Dimm  224  GCTTTTTGCCTATGACATAAGTGTCTGATTTGGAGTCGACACAA~~~~~~~~~~~~~~~~
Dgri  1    ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Sleb  329  ATATACATACATAGTATACGAGTATGTTGTATACCCACATGTTTGCGTTTCCGTTTCAAA
```

```
Dmel  221  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dsim  216  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dyak  215  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dana  293  CTATAACCTATTTTTGAACACCTATTTTGTTTCGGATCCGATCCTGTTGAAAAATCTGAA
Dpse  288  CTTAGGCAGGAATAAGTTAGTTAAAGATGTACAGAATTTGCATATTCGATTGGTTGAATA
Dper  295  CTTAGGCAGGAACAAGTAAGTTAAAGATGTACAGAATT~~~~~~~~~~~~~GGTTGAATA
Dwil  211  TTTTGTGATATTTCTTGGTATATAGATAGATAGCTGAAACAAAATGTGA~~~~~~~~~~~
Dhyd  1    ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dmoj  328  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dvir  333  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dimm  267  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dgri  1    ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Sleb  389  AAACCCTTTTCAGGCGCTCGTATCATACACTAGACGCCATAATGGCGTGGCGACATTTTG

Dmel  221  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dsim  216  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dyak  215  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dana  353  AATACCCCTTAATCAGCGCGTAGACTGGACTTAGCAATGAACAATGGGGAACTCTGCGGC
Dpse  348  GTTTACTGAGGGCAGAAAGGAAGATTCAACCTCATTTTGAGAGTATGAAAATCCTCTTTA
Dper  342  GTTTACTGAGAGCAGAAAGGAAGATTCAACCTCATTCTGAGAGTATGAAAATCCTCTTTA
Dwil  259  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dhyd  1    ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dmoj  328  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dvir  333  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dimm  267  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dgri  1    ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Sleb  449  TTGTCATGCGAGTGCTTGTGTGTGGCAAGATTTTATAATAATACGCAGTTTTTGACTTGA

Dmel  221  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dsim  216  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dyak  215  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dana  412  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dpse  408  AAATAGTTTAAAATATCTTAAAGTATCTGGAAATTTTGGATGAGGAAAATCTTAACGGGG
Dper  402  AAATATCTTAAAGTATCT~~~~~~~~~~GGAAATTTTGGATGAGGAAAATCTTAACGGGG
Dwil  259  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dhyd  1    ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dmoj  328  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dvir  333  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dimm  267  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dgri  1    ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Sleb  509  CTTTGGCAGCTTTAGGTTCTCGTCTTGGCTTGCGTGTTGCTCATCAGAGTAATGGCAAAG

Dmel  221  ~~~~~~~~~~~~TGGGAGCCAACGCCTTCTGCCTATCTGCCGCAGAACAGGCGAGAACGG
Dsim  216  ~~~~~~~~~~~~TGGGAGCCAACGCCTTCTGCCTGTCTGCCGCAGAACAGGCCAGAACGG
Dyak  215  ~~~~~~~~~~~~TGGGAGCCAACGCC~~~~~CTCGTCTGAGA~~~~ATTCACAGAAGCGG
Dana  412  ~~~~~~~~~~~~TGGGAGCCAACGCC~TCTGTCTGGTGGTAAG~GAAACAGCACTAAA~~
Dpse  468  AGCCCTTGATGTTGGGAGCCAACGCCT~~~~~CTGTCTG~~GCAGAAACAATAGATCGTC
Dper  452  AGCCCTTGATGTTGGGAGCCAACGCCT~~~~~CTGTCTG~~GCAGAAACAATAGATCGTC
Dwil  259  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dhyd  1    ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dmoj  328  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dvir  333  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dimm  267  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dgri  1    ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Sleb  569  TGAAATAGACATGACATAAGATCTGTTCAAAATCGTATGCATGATCTTCTCACACACACA
```

```
                                                  |--Dll304Min start-->
Dmel  270   ACAAAGGAG~~~~~~~~~~~~TCCCAATACCATTCCTGTGCCAATGGGATTATCTATGAG
Dsim  265   ACAAAGGAG~~~~~~~~~~~~TCCCAATCCCATTCCTGTGCCAATGGGATTATCTATGAG
Dyak  255   ACAAAAGAG~~~~~~~~~~~~TCCCAATCTCATTCCTGTGCCAATGGGATTATCTATGAG
Dana  456   ~~~~~~~~~~~~~~~~~~~~~T~CCAATCCCATTCCTATGCCAATGGGATTATCTATGAG
Dpse  521   CAGCCGCAATCCCATTCCCATTCCGATTCCCATTCAAATGCCAATGGGATTATCTATGAG
Dper  505   CAGCCGCAATCCCATTCCCATTCCGATTCCCATTCAAATGCCAATGGGATTATCTATGAG
Dwil  260   ACAAAAGATGCT~~~~~~~~~~~~~~~~~~CATTATAATGCCAATGGGATTATCTATGAG
Dhyd  1     ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dmoj  328   ~~~~~~~~~~~~~~~CACACAAAAGCGGCTCATTATAATGTCAATGGGATTAACTATGGA
Dvir  333   ~~~~~~~~~~~~~~~CACACAAAAGCGGCTCATTATAATGCCAATGGGATTAACCATGGA
Dimm  267   ~~~~~~~~~~~~~~~CACACAAAAGCGGCTCATTATAATGCCAATGGGATTATCTATGGC
Dgri  1     ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~TGCCAATGGGATTATCTATGGA
Sleb  629   CAGGTGTACAGAAAAGATACAAAAGAGTCTGATTGTAATGCCAATGGGATTATCTATGGG

Dmel  317   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dsim  312   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dyak  302   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dana  494   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dpse  580   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dper  564   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dwil  301   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dhyd  1     ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dmoj  374   CAGCACACAGACACACGAGCACACACAAACTCAAACACACACAATCATGCATACCCATGT
Dvir  379   CAACACACACACACACGCAGATACACACACATACACACACTCGTAGAGAGC~~~~~~~~~
Dimm  312   ~~~~~~~~~CAGACACGC~~~~~~~~~~~~~~~~~~~~~~~~~~AGAGAGACTTTCATAA
Dgri  23    CACCACACACACACACACACGCACACACATATATTTATATATATATATATATAGACAATC
Sleb  688   ~~~~~~~~~~~~~~~~~~~~~CCGCAGCTTGTCATACGCCACTCGCACTATCACTAGCAC

                                                           BTD
                                                    PAN  ------
                                                   ------
Dmel  317   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~CTGTGAAATCCTTTTAGGCGGAAACTATGG
Dsim  312   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~CTGTGAAATCCTTTTAGGCGGAAACTATGG
Dyak  302   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~CTGTGAAATCCTTTTAGGCGGAAACTATGG
Dana  494   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~CTGTGAAATCCTTTTAGGCGTAAACTATGG
Dpse  580   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~CTGTGAAATCCTTTTAGGCGGAAATTATGG
Dper  564   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~CTGTGAAATCCTTTTAGGCGGAAATTATGG
Dwil  301   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~CTGTGAAATCCTTTTAGGCGTAAATTATGA
Dhyd  1     ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~GTGAAATCCTTTTAGGCGGAAATTAT~G
Dmoj  434   TCGGCTAGACATTCAATAATG~CTGTGAAACTGTGAAATCCTTTTAAGCGGAAATTAT~G
Dvir  429   ~~~~~~GATGCCTGAAAA~~~~~~~~~~~~CTGTGAAATCCTTTTAGGCGGAAATTAT~G
Dimm  338   TGCCCTGAANCTGGAAATCTTAnnnnnnnn~~GTGAAATCCTTTTAGGCGGAAATTAT~G
Dgri  83    GCATTCATAATGC~~~~~~~~~CTGTGAA~CTGTGAAATCCTTTTAGGTGGAAATTATGA
Sleb  728   TA~~~~~~~~~~AGTCATAATGCTTGGAA~CTGTGAAATCCGTTTAAGCGGAAATTAT~G
```

```
                                       PAN
                                      ------
                                    Motif A      Motif A
                                    --------     --------
Dmel  348  AA~CCCACACACAGG~~~~~~~~~~~~~CACAAAGC~CAG~CACAATGC~~~~~~~~~~~
Dsim  343  AA~CCCACACACAGG~~~~~~~~~~~~~CACAAAGC~CAG~CACAATGC~~~~~~~~~~~
Dyak  333  AA~CCCACACACAGGCACAGCGCACAACGACAAAGC~CAG~CACAATGC~~~~~~~~~~~
Dana  525  AA~CCCACACACAGCTA~~~~~~~~~~~CAGGCTGCCCAGACACAATGC~~~~~~~~~~~
Dpse  611  AAACCCACA~~~~~~~~~~~~~~~~~~~CACAATGTGCAG~GACAATGC~~~~~~~~~~~
Dper  595  AAACCCACA~~~~~~~~~~~~~~~~~~~CACAAAGCGCAG~GACAATGC~~~~~~~~~~~
Dwil  332  AA~CCAAGGACCATTGTGGCCAAAGAGA~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dhyd  28   AAACGGCCG~ACAG~C~~~~~~~~~~~~~~AGGG~~~CA~~CACAATGC~~~~~~~AGGA
Dmoj  492  AAACGGTCG~ACAG~~~~~~~~~~~~~~~~~~~~~~CAGGATACAATGC~~~~~~~AGGA
Dvir  471  AAACGGCCA~ACAGGCG~~~~~~~~~~~~~AGG~~~~AG~~GACAATGC~~~~~~~~~~~
Dimm  395  AAACGCAGCAGGAG~C~~~~~~~~~~~~~~AGGACACAGA~CACAATGG~~~~~~~G~~~
Dgri  133  AA~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~CAG~GACAATGC~~~~~~~~~~~
Sleb  776  AAACACTGGCGAAGG~~~~~~~~~~~~~~~~~~ACACACAAATGGGAGGACAAAAAGCGC

                                   Motif B    Hox-like
                                   -------    --------
Dmel  380  ~~~AACAAGTGTTGCGGCAGATTGAGCAACAAAAGGCTCATAATTGTGGAAGC~~~~~~~
Dsim  375  ~~~AACAAGTGTTGCGGCAGATTGAGCAACAAAAGGCTCATAATTGTGGAAGC~~~~~~~
Dyak  378  ~~~AACAAGTGTTGCGGCAGATTGAGCAACAAAAGGCTCATAATTATGGAAGC~~~~~~~
Dana  561  ~~~AACAAGTGTTGCGGCAGATTGAGCAACAAAAGGTTCATAATTGTGTAAGC~~~~~~~
Dpse  639  ~~~AACAAGTGTTGCAGCAGATTGAGCAACAAAAGGTTCATAATTATGGAAGC~~~~~~~
Dper  623  ~~~AACAAGTGTTGCAGCAGATTGAGCAACAAAAGGTTCATAATTATGGAAGC~~~~~~~
Dwil  358  ~~~AACAAGTGTTACAACAGATTGAGCAACAAAAGGACTATAATTGTTAAAGC~~~~~~~
Dhyd  60   GGACACAAGTGTTGCGGCAGATTGAGCAACAAAAGAGCAATAATTGTTAGAGCGAGAG~~
Dmoj  522  GGACACAAGTGTTGCGGCAGATTGAGCAACAAAAGGGCAATAATTGTTAGAGCGAGAG~~
Dvir  499  ~~GCACAAGTGTTGCGGCAGATTGAGCAACAAAAGGGCCATAATTGTTAG~~~~~~~~~~
Dimm  428  ~CGCACAAGTGTTGCGGCAGATTGAGCAACAAAAGGGCCATAATTGTCAGTGC~~~~~~~
Dgri  145  ~CGCACAAGTGTTGCGGCAGATTGAGCAACAAAAGGGCCATAATTGTTAGTGCCTCTTAA
Sleb  818  GCACACAAGTGTGGCGGCAGATTGAGCAACAAAAGGGCCATAATTGTTCGA~~~~~~~~~

                        Motif A                             BTD
                        --------                            ----......
Dmel  430  ~~~~~~~~~~~CACACAATGC~~~~~~GACACATTC~~~~~~~~~~GAGAGCGG~~~~~~
Dsim  425  ~~~~~~~~~~~CACACAATGC~~~~~~GACACATTC~~~~~~~~~~GAGAGCGG~~~~~~
Dyak  428  ~~~~~~~~~~~CACACAATGC~~~~~~GACACATTC~~~~~~~~~~GAGAGCGA~~~~~~
Dana  611  ~~~~~~~~~~~CACACAATGC~~~~~~GACACATTG~~~~~~~~~~GAGGGCGAGAGCGA
Dpse  689  ~~~~~~~~~~~CACACAATGC~~~~~~GACACATTCGATG~~~~~CGAGACCGGGGCGAG
Dper  673  ~~~~~~~~~~~CACACAATGC~~~~~~GACACATTCGATG~~~~~CGAGACCGGGGCGAG
Dwil  408  ~~~~~~~~~~~CACACAAGGA~~~~~~CTGACCCTCCCTATCTCACTGTGTGTGTGTATG
Dhyd  117  ~~~~~~~~~~~ACAACAATGC~~~~~~GACACATTC~GTG~~~~CAACAGGCCAAAGGCA
Dmoj  579  ~~~~~~~~~~~~~~ACAATGC~~~~~~GACACATTTCCTG~~~~CAACAGGCCAAAGGCA
Dvir  547  ~~~~~~~~~~~~~~ACAATGC~~~~~~GACACATTC~GTG~~~~CAACATTC~~~~~~~~
Dimm  480  ~~~~~~~~~~~CACACAATGC~~~~~~GACACATTGC~~~~~~~~~~~~~~~~CGCTGCA
Dgri  204  ~~~~~~~~~~~CACACAATGC~~~~~~GACACAGCCACAAATCTGTAGGTGTGGCTGCCA
Sleb  868  ~~~~~~~~~GCCACACAATGCGCCTTTGACACATTG~~~~~~~~~~~~~~~~TGTGAGGG
```

```
              ..BTD
                 --
Dmel  457  ~~~~~~GGATGAGGACGAGTCCAG~GGGACTGC~~~CGGTCCTTCGTTGTTCTCCATGG~
Dsim  452  ~~~~~~GGATGAGGACGAGGCCAG~GG~ACTGC~~~CGGTCCTTCGTTATGGTCCATGG~
Dyak  455  ~~~~~~GGATGAGGATAGTC~~~~~~~~~~~~~~~~~GGTCCTTCGTTATGGTCTATGGT
Dana  645  TTGCG~GTTGAAGGGAAAAGGACAAAGACTAG~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dpse  728  GACGA~GGATGTGGACGAGGGACGAGGGACTGT~~~AGCAACC~~~~~~~~CTCGCTGG~
Dper  712  GACGA~GGATGTGGACGAGGGACGAGGGACTGT~~~AGCAACC~~~~~~~~CTCGCTGG~
Dwil  452  TGTGTGTGTGTGTGGGTTTGGTCACAGGTCATA~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dhyd  156  ACAGG~GGGCGAACCACAAATCCTGCCAAGC~~~TGTAGGTGTGGCTGCCAGCCAACGGG
Dmoj  616  ACATGTGGGCGAACCACAAATCCAGCCAAGCACCAATAGGTGTGGCTGCCAGCCAACAAT
Dvir  574  ~~~GG~AGGCGAACCACAAAACCCGACAGGC~~~TGTAGGTGTGCCCGATGG~~~~~~~~
Dimm  508  ACAGA~AGGCGAGTCCGAG~TCC~GA~AAGT~~~TGTAGGTGTGGCTGATG~~~~~~~~~
Dgri  248  GCAACAATGTACTCTATCTCAGTATCTCTTTCTCTCTCTCGCTCTCTCTTTCACTCATGC
Sleb  904  AAGGAACACTGAGTCCG~~~~~~~~~~~~~~~~~TGTAGGTGTGGCCATCCAGACTTATC

Dmel  507  CAT~GGTACTCGGTAAT~~~~~~~~~~~~~~~~~~~~~~~GTAG~~~~~~~~~~~~~~~~
Dsim  501  CAT~GGTACTCGGTAGT~~~~~~~~~~~~~~~~~~~~~~~GTAG~~~~~~~~~~~~~~~~
Dyak  493  T~~~GGTCTA~~~TAGTTGGTA~~~~~~~~~~~~~~~~~~GTAG~~~~~~~~~~~~~~~~
Dana  675  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~G~~~~~~~~~~~~~~~~
Dpse  775  CAGAGGTTTGTTCGGGG~~~~~~~~~~~~~~~~~~~~~~~TTAGGGCCAA~~~~~~~~~~
Dper  759  CAGAGGTTTGTTCGGGG~~~~~~~~~~~~~~~~~~~~~~~TTAGGGCCAA~~~~~~~~~~
Dwil  484  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dhyd  212  C~~~~CAACAATGCCAC~~~~~~~~~~~~~~~~~~~~~~~TCGCATGTCCTTTTCCTTGC
Dmoj  676  C~~~~CAACAATGCGCC~~~~~TCTTATGTCCTTTGGCTTTCTTTCTTTCTTT~CTTTCT
Dvir  619  ~~~~~CAACAATGCCCG~~~~~~~~~~~~~~~~~~~~~~~TTTTGTTCGTTTATCAATTC
Dimm  551  ~~~~~CAACAATGCCGC~~~~~~~~~~~~~~~~~~~~~~~AATAGGATCGCGGCTAGAGG
Dgri  308  TCTCGCAGTCCGTTCCTTTGTGGACCAA~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Sleb  947  CAAGATACGAGTATAGGCGA~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

Dmel  526  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~GGCTGCTGACCAGGCTAATGAG~T
Dsim  520  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~GGCTGCTGACCAGGCTAATGAG~T
Dyak  512  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~GTCTGCTGGCCAGGCTAATGAG~T
Dana  676  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~GTCTGTTGGCCTGGCTAATGAG~T
Dpse  801  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~GGCAGTTGGCCAGGCTAATGAG~T
Dper  785  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~GGCAGTTGGCCAGGCTAATGAG~T
Dwil  484  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~GGTCACTAATGAGAG
Dhyd  245  TTTCTTTCT~TTCCATTTCCAATGCTATTTTATTTT~~~~~~~~~~~~~~TTAATGAGCT
Dmoj  726  TTTCTTCCACTTCCAATTCCATTTCTATTTTATTTT~~~~~~~~~~~~~~TTAATGAACT
Dvir  652  TTTTGCGCTGCG~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~CCAG~CTAATGAGCT
Dimm  584  CTAGTGGCAACCTCTTGTTCCATGGCCT~~~~~~~~~~~~~~~~~~~~~~GTAATGAGCT
Dgri  335  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~CTAATGAGCG
Sleb  966  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~AGGCTAATGAGCA
```

```
                                         Motif B
                                         -------
Dmel  550  G~~~~~~~~~G~~~~~~~~~~~~TCGC~~AAACAAACCATAAA~~~~~~~~~~~~~~~~~
Dsim  544  G~~~~~~~~~G~~~~~~~~~~~~TCGC~~AAACAAACCATAAA~~~~~~~~~~~~~~~~~
Dyak  536  G~~~~~~~~~G~~~~~~~~~~~~CCGC~~AAACAAACCATAAA~~~~~~~~~~~~~~~~~
Dana  700  G~~~~~~~~~G~~~~~~~~~~~~CCGA~~AAACAAACCATAAGTTAGCCCAAATCCTGCA
Dpse  825  G~~~~~~~~~G~~~~~~~~~~~~CCGCA~AAACAAACCATAAA~~~~~~~~~~~~~~~~~
Dper  809  G~~~~~~~~~G~~~~~~~~~~~~CCGCA~AAACAAACCATAAA~~~~~~~~~~~~~~~~~
Dwil  500  TAGAGGAGATGGCATA~~~~~~~ACGAA~AAACAAACCATAACCATAAAATTCCATCACA
Dhyd  290  G~~~~~~~~~~~~~~~~~~~~~~CAGCAAAAACAAACCATAAA~~~~~~~~~~~~~~~~~
Dmoj  772  G~~~~~~~~~~~~~~~~~~~~~~CAGCAAAAACAAACCATAAA~~~~~~~~~~~~~~~~~
Dvir  678  G~~~~~~~~~~~~~~~~~~~~~~CAGCAAAAACAAACCATAAA~~~~~~~~~~~~~~~~~
Dimm  622  GCCGCTGCT~G~~~~~~~~~~~~CAGCAAAAACAAACCATAAA~~~~~~~~~~~~~~~~~
Dgri  346  ACAG~~~~~~~~~~~~~~~~~~~~CGAAAA~~CAAACCATAAAAGATGTAAAAAGGTGCC
Sleb  980  GAGGTCGCTTG~~~~~~~~~~~ACCGCAAAA~CAAACCATAAA~~~~~~~~~~~~~~~~~

Dmel  570  TGCC~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~GATTTCGTGCGG~~T~CCA~~~~~~~
Dsim  564  TGCC~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~GATTTCGTGCGG~~T~CCA~~~~~~~
Dyak  556  TGCC~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~GATTTCGTGCGG~~T~CCA~~~~~~~
Dana  737  CGCC~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~GATTT~~~~~~~~~~~~~~~~~~~~~
Dpse  846  CCATTGGAAATCTTGCACA~~~~~~~~~~~~~TTGATTTCATGCCC~~T~CCGACTTT~~
Dper  830  CCATTGGAAATCTTGCACA~~~~~~~~~~~~~TTGATTTCATGCCC~~T~CCGACTTT~~
Dwil  552  TT~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dhyd  311  AATTATCT~~~~~~~~~~~~~~~~~~~~~~AAAA~TGT~~~~~~~~~~TGCCACAT~~TG
Dmoj  793  AATTATCT~~~~~~~~~~~~~~~~~~~~~~AAAA~TGT~~~~~~~~~~TGCCACAT~~TG
Dvir  699  AATTGTCCGAACATTTT~~~~~GTGGCCACAATGGTGTCTAACCAAAATGCCACATGCTG
Dimm  652  CTGGAATCGAATAAAGAAAAAGGTG~CCACAATTGATTTCATGCGGC~TGCCAC~~~~~~
Dgri  384  ACATTGCATCTCATGCAA~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~CCT~~~~~~~
Sleb  1010 ~~~~~~~~~~~~~~~~~~~AATGTG~CCGCTTC~GATTTCATTTTCATGCAATCTGTTGT

Dmel  589  ~~~~~~~~~~~~~~~~~GCCT~~~~CGGCGGCAAC~~~TTCTTTTAATACC~AC~~~~~~
Dsim  583  ~~~~~~~~~~~~~~~~~GCCT~~~~CGGCGGCAAC~~~TTCTTTTAATACC~AC~~~~~~
Dyak  575  ~~~~~~~~~~~~~~~~~GCCT~~~~CGGCGGCAAC~~~TTCTTTTAATTCC~AC~~~~~~
Dana  745  ~~~~~~~~~~~~~~~~~~~~~~~~~~GGCGGCATC~~~~~CTTTTTATTGC~CTC~~~~~
Dpse  887  ~~~~~~~~~~~~~~~~~~~~~~~~~~GGCGGCATCTTGTTCCTTCTCTACC~AC~~~~~~
Dper  871  ~~~~~~~~~~~~~~~~~~~~~~~~~~GGCGGCATCTTGTTCCTTCTCTACC~AC~~~~~~
Dwil  553  ~~~~~~~~~~~~~~~~~~~~~~~~~~GGCGGCA~~~~~TTCAAGCATCTGGCCATTTGAT
Dhyd  336  TT~~~~~~~~~~~~~~~GCCG~~~~CCGCGGCAACT~~TTCCTTTGTGGCC~ACCTACAA
Dmoj  818  TTGAATGCATGCAACATGCCGC~CGCCGCGGCAACT~~TTCCTTTGTGGCC~ACCTACAA
Dvir  754  T~~~~~~~~~~~~~~~~GCCGAT~~~GGCGGCGACT~~TTCTTTTGAGCCC~ACCTACAA
Dimm  703  ~~~~~~~~~~~~~~~~~~~~~~~~~~GGCGGCAACT~~TTGCTTTGAGCCC~ACCTACAA
Dgri  404  ~~~~~~~~~~~~~~~~~GCTG~~~~CGGCGGCAAAGA~TTCTTTTGAGCCCACCTACAAC
Sleb  1050 GTGGGACTCT~~~~~~~~~~~~~~~~GGCGGTGAAT~~TTTCTTTGAGGCCCACCTACAA

Dmel  618  ~~~~~~~~~~~~~~~~~~~~~~~TTGACACTTGGTCAAGATCTAGGA~~~TACCCA~~TT
Dsim  612  ~~~~~~~~~~~~~~~~~~~~~~~TTGACACTTGGCCAAGATCTAGGA~~~TACCCA~~TT
Dyak  604  ~~~~~~~~~~~~~~~~~~~~~~~TTGACACTTGGCCAAGATCTAGGA~TCTAGGATCGCC
Dana  768  ~~~~~~~~~~~~~~~~~~~~~~~TTGACACTTGGCTCGATGAGGCCT~ATGGGCTATGAG
Dpse  914  ~~~~~~~~~~~~~~~~~~~~~~~TTGACACCCTGCTTTTGGGTCAGA~TTTTCCCATTTT
Dper  898  ~~~~~~~~~~~~~~~~~~~~~~~TTGACACCCTGCTTTTGGGTCAGAATTTTCCCATTTT
Dwil  583  GTGGCCACCAACACAATCGTTG~TTGACACTTGTTGAGATTGCATTGTACTATTCAGGTA
Dhyd  374  CGCTGCGCTC~TTGTTGCTGCTGTTGACACTTGGT~~~~~~~~~~~~~~~~~~~~~~~~~
Dmoj  874  CGCTCC~~~~~~~~~~~~~GTTGTTGACACTTGGT~~~~~~~~~~~~~~~~~~~~~~~~~
Dvir  792  CGC~~~~~~~~~~~~~~~~GCTGTTGACACTTGCTGCTGCCGCCGCG~TCCTGCG~~~TT
Dimm  735  CGTA~~~~~~~~~~~~~~TGTTGTTGACACTTGTTGCAGCCGTGG~~~~~~~~~~~~~~~
Dgri  443  GTCTG~~~~~~~~~~~~~~~~~~TTGACACTTGTGGCAGGCAGCCCAGCTAACTCA~~~~
Sleb  1092 CGCTGTTGTTGTTGTTGCCGCTGTTGACACTCTTCGCAGCGGTTCGA~GGTAAAGGCCAT
```

```
Dmel  651  CCACTAGGCA~~~~~~~~~~~~~~~~~T~GGAATTTATGTCCG~~~~~~~~~~~~~~~~~
Dsim  645  CCACCAGACA~~~~~~~~~~~~~~~~~T~GGAATTTATGTTCG~~~~~~~~~~~~~~~~~
Dyak  641  GATCGCCCCATTCCCATCCACTCGCAGT~GGAATTTATGTCCG~~~~~~~~~~~~~~~~~
Dana  805  AGGCCCGAACTTATCCAATC~~~~~~~T~CGAATTTATGTTCG~~~~~~~~~~~~~~~~~
Dpse  951  CCCCGTAACA~~~~~~~~~~~~~~~~~AAAGAATTTATGTTCG~~~~~~~~~~~~~~~~~
Dper  936  CCCCGTCGCCTCT~~~~~~~~~~~~~~~~~GAATTTATGTTCG~~~~~~~~~~~~~~~~~
Dwil  642  TATATATATACGTATATAT~~~~~~~~~~~AGATTTATGTTGG~~~~~~~~~~~~~~~~~
Dhyd  407  ~~~~~~~~~CA~~~~~~~~~~~~~~~~~~~AGATTTATGCCCCA~CTCCAC~~~~~~~~~
Dmoj  895  ~~~~~~~~~CA~~~~~~~~~~~~~~~~~~~AGATTTATGCCCCA~CTCCACCAACACCAA
Dvir  832  CCT~~~~~~GA~~~~~~~~~~~~~~~~~~~AGATTTATGCGCTCACTCCCCCCCCCCA~~
Dimm  765  ~~~~~~~~~CA~~~~~~~~~~~~~~~~~~~AGATTTATGCCCCA~~~~~~~~~~~~~~~~
Dgri  480  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~AGATTTATGCCCCCAAA~~~~~~~~~~~~~
Sleb  1151 TGCAC~~~~~~~~~~~~~~~~~~~~~~~~~AGATTTATGTTCG~~~~~~~~~~~~~~~~~

                                       Hox/EXD
                                       ---------
                          MATAYTTGS~GMAAWTAAAT            HTH
                          --------------------           -----
Dmel  675  ~~~~~~~~~~~~~ACAATATTTGG~GAAATTAAATCATT~CCCGCGGACAGTTTTATAGT
Dsim  669  ~~~~~~~~~~~~~ACAATATTTGG~GAAATTAAATCATT~CCCGCGGACAGTTTTATAGT
Dyak  682  ~~~~~~~~~~~~~ACAATATTTGG~GAAATTAAATCATT~GCCGCGGACAGTTTTATAGT
Dana  839  ~~~~~~~~~~~~~ACAATATTTGG~GAAATTAAATCATT~GCCGCCGACAGTTTTATGTT
Dpse  976  ~~~~~~~~~~~~~AAAATATTTGG~GAAATTAAATCATT~GCCGCCGACAGTTTTATAGT
Dper  961  ~~~~~~~~~~~~~AAAATATTTGG~GAAATTAAATCATT~GCCGCCGACAGTTTTATAGT
Dwil  673  ~~~~~~~~~~~~~AAAATATTTGG~GAAATTAAATCATTTGAAGTGGACAGTTTTATGGT
Dhyd  429  ~~~~CGAAAAAAAAACATATTTGG~GAAATTAAATCATT~GGCGCGGACAGTTTTATAGA
Dmoj  927  ACAACGAAAAA~~TATATATTTGCGGAAATTAAATCATT~GACGCGGACAGTTTTATAGA
Dvir  864  ~~~~~~~~~~~~~AACATATTTGG~GAAATTAAATCATT~GGCGCGGACAGTTTTATACG
Dimm  781  ~~~~~~~~~~~~~ACAATATTTGG~GAAATTAAATCATT~GGCGCAGACAGTTTTATAGT
Dgri  497  ~~~~~~~~~~~~~AC~ATATTTGG~GAAATTAAATCATT~GCCGCGGACAGTTTTATGGG
Sleb  1168 ~~~~~~~~~~~~~AAAATATTTGG~GAAATTAAATCATT~GGCGCGGACAGTTTTATAGG

           <-Dll304Min-|
Dmel  721  GC~~GGGCGTGGCTGGTATTGGA~~~~~~~~~~~~~~~GGAGGGAGGATG~~~~GAGGAT
Dsim  715  GC~~GGGCGTGGCGGGTATTGGA~~~~~~~~~~~~~~~GGAGGGAGGACG~~~~G~~~~~
Dyak  728  GC~~GGGCGTGACGGGAC~~~~~~~~~~~~~~~~~~~~GGGAGAAGGGGATTGGGATGGG
Dana  885  CCA~GGGCGTGACGAGGGAACCGTGCTGGGGAAAACTATTCGGGGAAGCTATGTGGGGAA
Dpse  1022 CCA~GGGCGTGACGG~~~~~~~~~~~~~~~~~~~~~~~~~AGGGAATC~~~~~~GAGGAG
Dper  1007 CCA~GGGCGTGACGG~~~~~~~~~~~~~~~~~~~~~~~~~AGGGAATC~~~~~~GAGGAG
Dwil  720  TTTAGGGCGTGACAAACGTTACCAATGATTAGATAGGGGTTTA~~~~~~~~~~~~~~GAT
Dhyd  484  CA~~GGGCGTGACGGCGGCGGCGGCAGCGGCGGGCGT~GGAGGATG~~~~GCTAGCTGGT
Dmoj  984  CG~~GGGCGTGACGGCGGCGGGCGTACAGGTCGATGCAACGGCAAGGGGCAACGAGCCGG
Dvir  910  CG~~GGGCGTGACGGCAACA~~~GCAGGGGCACAGCTAGGAG~~TG~~~~GGAG~~~~~~
Dimm  827  CC~~GGGCGTGACGACATTGTGGCT~~~~~~~~~~~~~GGAGCAACCAGAGGAGCGGAGA
Dgri  542  CCAGGCCCGTGACGGTGGCGGCAACGGCAGCGGTCGCTGGTTCTAAGCGGGAATCTCTAC
Sleb  1214 CC~~GGGCGTGACGG~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
```

```
Dmel  760  GGTGGATGGTGGATGGAGGGAGGGT~~~~~~~~TCGTGCTGGGGAAGGGGATGGG~~~~~
Dsim  748  ~~~~~~~~~~~~~TGGAGGGAGGGT~~~~~~~~TCGTGCTGGGGAAGGGGATGGG~~~~~
Dyak  766  A~~~~~~~~~~~~~AAGGGGAAGGGGATA~~~~~~~~~~~~~~~~~~GGGGCTGG~~~~~
Dana  944  ACAATGAGCACCACCCGGAGGTTCTCCTGCAGCAGATAGCTCCGTGAGGTCCGGGTCCGG
Dpse  1049 ~~~~~~~~~~~~~~~~~~~~T~~~~~~~~~~~~TCGTG~TGTT~~~GGGGGTAAAA~~~~
Dper  1034 ~~~~~~~~~~~~~~~~~~~~T~~~~~~~~~~~~TCGTG~TGTT~~~GGGGGTAAAA~~~~
Dwil  766  GGAATGTGAGGGATGCGGGGAGGGC~~~~~~~~GAAAGGTGGGGAAAGGGGCACAGATTT
Dhyd  537  TGCTGCAACGGGTC~GGGGGCATTGGTAGGGG~~~CTGCTGGATCTAAGCGGGCAGCGCT
Dmoj  1042 GT~~~~~~CGGTTC~GGGGGCATGGGTAAGGG~~~CTGCTGGATCTAAGAGAGCATCTGT
Dvir  952  ~~~~~~~~CGGGGCCGGGGGCAGGGGCCGGGGTTGCTGCTGGATCTAAGCGAGCTGCG~~
Dimm  872  GA~~~~~~~~~~~~~~~~~~~~~~~~~~~GGGTAGCTGCTGGATCTAAACGGGAATGGCC
Dgri  602  TATAGCCTATAATAGCTATAACTGTGGCCTCCGTGTAGGAGTTGAACAATGCCCGTT~~~
Sleb  1226 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

Dmel  806  ~~~~~~~~~CTATCTA~~~~~ACAGTGACCTCAG~~~~~~~~~~~~~~~~~~~~~~~~~~
Dsim  782  ~~~~~~~~~CTATCAA~~~~~ACAGTGACCTCAG~~~~~~~~~~~~~~~~~~~~~~~~~~
Dyak  789  ~~~~~~~~~CTATCTA~~~~~ACAGTGACCTCAGCCCCCGCTGAACCCACGAG~~~~~~~
Dana  1004 GCAAATAGCACGAAC~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dpse  1069 ~~~~~~~~~CTATCTA~~~~~ACAATGACCTCTCCCTCTCTGGGCTCTCT~~~TCCCCTC
Dper  1054 ~~~~~~~~~CTATCTA~~~~~ACAATGACCTCTCCCTCTCTGGGCTCTCT~~~TCCCCCC
Dwil  818  CCACATGAAAATCTAACAAAAGCCTTTGCCG~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dhyd  593  GACTCTGACTCTGTCGGC~~~~~~~~~GACTGTGG~CCTCCGTGTGGGCG~~~CTGAACA
Dmoj  1092 GACTCTGACTC~~~~~~~~~~~~~~~~~~~TGTGGCACTCCGTGTGGGCGCCTCTGAACA
Dvir  1002 ~ACTTTGGCTC~~~~~~~~~~~~~~~~~~~TG~G~~CCTCCGTGTGGGCG~~~CCGAACA
Dimm  905  AGCAATGGGCTATGAATTTACAACTGTGAGTGTG~CCCTCCGTGTGGCAG~~~CCAAACA
Dgri  658  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Sleb  1226 ~~~~~~~~~~~~~~~~~TCATAGGCGGGAGGCAGCGAGACAGCGAGACAGCGATCCAACA

Dmel  826  ~~~~~~~CCCCC~~GCTGAATC~~CACGAGTGGGAAAATTGGA~~~~~~~~~~~~~~~~~
Dsim  802  ~~~~~~~CCCCC~~GCTGAACC~~CACGAGTGGGAAAATTGGA~~~~~~~~~~~~~~~~~
Dyak  828  ~~~~~~~C~~~~~~~~~~~~~~~~~~~~~~~GGGAAAATTGGA~~~~~~~~~~~~~~~~~
Dana  1018 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~GAGCAAATTGGA~~~~~~~~~~~~~~~~~
Dpse  1113 CCCCCCCCCCCC~~GCACTTCCAGCACGAATGAGCAAATTGGA~~~~~~~~~~~~~~~~~
Dper  1098 CCCCCCCCCCCCCCGCACTTCCAGCACGAATGAGCAAATTGGA~~~~~~~~~~~~~~~~~
Dwil  848  ~~~~~~~~~~~~~~~~~~~~~~~~~ACGACTGAGCAAATTGGAG~~GGGCAAAAAAATGA
Dhyd  640  ATGCCCGTT~~~~~~~~~~~~~~~~~~~~~~GAGCAA~TTGGA~~~~~~~~~~~~~~~~~
Dmoj  1133 ATGCCCGTT~~~~~~~~~~~~~~~~~~~~~~GAGCAA~TTGGA~~~~~~~~~~~~~~~~~
Dvir  1037 ATGCCCGTT~~~~~~~~~~~~~~~~~~~~~~GAGCAA~TTGGGGA~~~~~~~~~~~~~~~
Dimm  961  ATGCCTGTT~~~~~~~~~~~~~~~~~~~~~~GAGCAA~TTGGAA~~~~~~~~~~~~~~~~
Dgri  658  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~GAGCAA~TTGGGTAA~~~~~~~~~~~~~~
Sleb  1270 ATGCC~~~~~~~~~~~~~~~~~~~~~~~AGCGAGCAA~TTGGA~~~~~~~~~~~~~~~~~
```

```
                    <--Dll304 End--|
                       Hox-like
                       --------
                    AATTGACA         Motif B     AATTGACA
                    --------         -------     --------
                                       Palindrome
                                       -----------
Dmel  858  ~~~~~AAAATGTCAATTATG~CGCAATTTTGCTTAAGCAATTGACA~TTTGTTGCTGCTC
Dsim  834  ~~~~~AAAATGTCAATTATG~CGCAATTTTGCTTAAGCAATTGGCA~TTCGTTGCTGCTC
Dyak  841  ~~~~~AAAATGTCAATTATG~CGCAATTTTGCTTAAGCAATTGGCA~TTTGTTGCTGCTT
Dana  1030 ~~~~~AAAATGTCAATTATG~CGCAATTTTGCTTGAGCAATTGACA~TTTGTGATTTGTG
Dpse  1153 ~~~~~AAAATGTCAATTATG~CGCAATTTTGCCTAAGCAATTGACA~TTTGTTGCCGCTG
Dper  1140 ~~~~~AAAATGTCAATTATG~CGCAATTTTGCCTAAGCAATTGACA~TTTGTTGCCGCTG
Dwil  882  GAAA~AAAATGTCAATTATG~CGCAATTT~GCTTAAGCAATTGACA
Dhyd  659  ~~~~~AAAAAGTCAATTATG~CGCAATTT~GCTTAAGCAATTGACATTTTGTTGCTGC~~
Dmoj  1152 ~~~~~AAAAAGTCAATTATAGCACAATTT~GCTTAAGCAATTGACATTTTGTTGTTGTTG
Dvir  1058 ~~~~AAAAAAGTCAATTATG~CGCAATTT~GCTTAAGCAATTGACATTTTTT~ACTCTTA
Dimm  981  ~~~~~AAAATGTCAATTATG~CGCAATTT~GCTTAAGCAATTGACAATTTGTT~~~~~~~
Dgri  672  ~~~~AAAAAAGTCAATTATG~CGCAATTT~GCTTAAGCAATTGACATTTTGTTTGTTTGT
Sleb  1288 ~~~~~AAAATGTCAATTATG~CGCAATTT~GCTTAAGCAATTGACATATTGTAGTAGCCC

Dmel  912  TG~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~TTGTTTGTTGTGATTGCA
Dsim  888  TG~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~TTGTTTGTTGTGATTGCA
Dyak  895  TG~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~TTGCGTTGTTTGTTGTGATTGCA
Dana  1084 CTGTGTTGCT~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~GTTTTGTTAGTGATTGCA
Dpse  1207 TGGCTGTTCGCTTTGTTGTT~~~~~~~~~~~~~~~~~~~~TGCTGCTTGTTGCGATTGCA
Dper  1194 TGGCTGTGGCTCTGACTGTTCGCTTTGTT~~~~~~~~GTTTGCTGCTTGTTGCGATTGCA
Dwil  939  TTTTATTTTTTCCATT~~~~~~~~~~~~~~~~~~GTTGTTGTTGCTATTTTCTGATTGCA
Dhyd  710  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~TGTTGT~~~~AGCTGGTTGCA
Dmoj  1207 T~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~AGCTGGTAGCA
Dvir  1112 TTGTTGTTGTTGTTGCTTTT~~~~~~~~~~~GTTGTTGTTGTTGTTGT~TGCTCGTTGCA
Dimm  1027 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~GTTGCTGTTGC~TATTG~TTGCA
Dgri  727  TGTTGTTGTTGCTCGTTGCCCGTTGCCCG~~~~~~~~~~~~~~~~~~~~~~~~~~TTGCA
Sleb  1342 ATGTTGTTGTT~~~~~~~~~~~~~~~~~~~~GTTGTTGTTTC~AGTTTTTTAAC~~TGCA

             MATAYTTGSGMAAWTAAAT
             -------------------
                             AATTGACA
                             --------
Dmel  932  CGCATACTTGGGCAAATAAATTGACAAATAGGATTTGCGT
Dsim  908  CGCATACTTGGGCAAATAAATTGACAAATAGGATTTGCGT
Dyak  920  CGCATACTTGGGCAAATAAATTGACAAATAGGATTTGCGT
Dana  1112 CGCATACTTGCGCAAACAAATTGACAAATAGGATTTGCGT
Dpse  1247 CGCATACTTGCTCAAATAAATTGACAAATAGGATTTGCGT
Dper  1246 CGCATACTTGCTCAAATAAATTGACAAATAGGATTTGCGT
Dwil  981  CGCATACTTGCTCAAATAAATTGACAAATAGGATTTGCGT
Dhyd  728  CGCATACTTGCGCAAATAAATTGACAAATAGGATTTGCGT
Dmoj  1219 CGCATACTTGCGCAAATAAATTGACAAATACGATTTGCAT
Dvir  1160 CGCATACTTGCGCAAATAAATTGACAAATAGGATTTGCGT
Dimm  1049 CGCATACTTGCGCAAATAAATTGACAAATAGGATTTGCGT
Dgri  761  CGCATACTTGGACAAATAAATTGACAAATAGGATTTGCGT
Sleb  1379 CGCATACTTGCGCAAATAAATTGACAAATAGGATTTGCGT
```

Ddc-.47

```
                                                             AFB
                                                            --------
Dmel   1    ~~~~~~~~~~~~~~~~~~~~~~~a~gaaaaa~~ccctgtttcga~~~~gtgactcataa~
Dsim   1    ttttttcttctgggagccc~~~~aagaaaaa~~ccctgtttcga~~~~gtgactcattt~
Dsec   1    ttttttcttcggggagccc~~~~aagaaaaa~~ccctgtttcga~~~~gtgactcatga~
Dyak   1    att~~~~~~~~~~~~~~~~~~~~~~~caaaa~~ccctgtttcga~~~~gtgactcatga~
Dere   1    ttttttctttggggagccc~~~~gacaaaaa~~ccctgttacga~~~~gtgactcatga~
Dpse   1    ~~~~~~~~~~~~~~~~~tt~~~~caaaaaatcgc~tgcccggtgtgacgtgactcatga~
Dper   1    agctcatgtgg~~~~~~tg~~~~gattcaaat~cgctgccggtgtgacgtgactcatga~
Dana   1    ~~~~~~~~~~~~~~~~~~~~~~~~agaaaatc~~~~~cctgg~attttgtgactcatga~
Dvir   1    ~~~~at~ctct~tgtcttgcctgaggaaa~~~~cccca~cgcagcgc~gtgagtcatgag
Dmoj   1    ~~~atcgc~ctctgt~tagtttgaggaat~~~~ccccatcacaacg~gatgagtcac~ag
Dgri   1    ~~~~~~~~~~~~~~~~t~~~~ttaggaaa~~~~cccca~tgcaacat~gtgactcataag

                     ETS
                   ---~----                   ETS                ETS
                           |--DdcΔ1-->     ---~~~~~--         ---~~~~~~-
Dmel   29   ~~ttggggga~ttcctgacgagatcgctctcttt~~~~~ccacaaattcgagt~~~~~~t
Dsim   49   ~~~tggggga~ttcctgtcgagatttttttt~~~~~~~~~tttaaatttgagt~~~~~~t
Dsec   49   ~~~tggggga~ttcctgtcgagatcgctctcttttttttcttttaatttgagt~~~~~~t
Dyak   30   ~~~tggggga~ttcccgacgagatcgctctctttt~~~~~cttaaattcgagt~~~~~~t
Dere   49   ~~~tggggga~ttcctgacgagatcgcgctctttt~~~~~cttaaattcgcgt~~~~~~t
Dpse   37   ~~~cggggga~ttcctggaccgcactc~~gcacggtc~cc~cataaaacg~~~~atacct
Dper   48   ~~~cggggga~ttcctggaccgcactc~~gcacggtc~cc~cataaaacg~~~~atacct
Dana   29   ~~~tggggga~ttcctgga~ggcacttatgtac~~acacctct~~~~~ctcttg~t~~~t
Dvir   49   ctcgggggaattcttgtcttga~~ctg~tcgatcc~~~a~~~~~actga~at~~~~~~~
Dmoj   49   ~~~tgagggaattgttttattta~~ctg~ttaaaaa~~~ac~tgga~ttatacgcaagct
Dgri   35   cgctgggggaattctttt~ttaatgcttcttaaaaatcaacct~aatt~at~~~tatgc~

            ---
Dmel   76   gggaagc~~~~~~~~~~~~~~~~~~~~~acgtgagtagaattcaaaatgttttgcttgct
Dsim   91   gggaagc~~~~~~~~~~~~~~~~~~~~~acgtgagcagcattcaaaatgtttggcttggt
Dsec   100  gggaagc~~~~~~~~~~~~~~~~~~~~~acgtgagcagcattcaaaatgttttgcttggt
Dyak   76   gagatgc~~~~~~~~~~~~~~~~~~~~~acgtgagcagaattcaaaacattccgctttgt
Dere   95   gggaagc~~~~~~~~~~~~~~~~~~~~~acgtgagcagaattcaaactgattcactaggt
Dpse   86   gggaa~~~cg~gcgc~~~~~ttggaggatcgtgcgcagaattcaaaccattctgagagg~
Dper   97   gggaa~~~cg~gcgc~~~~~ttggaggatcgtgcgcagaattcaaaccattctgagagg~
Dana   74   tggagaaacaagatcaggcttcgcacgaacgtgagaagaattcttcac~~~~~gatcgg~
Dvir   89   ~~~~~agtta~agagctg~~~~g~~~~~~~~~~~tt~g~~~~~at~t~~~att~t~c~~~
Dmoj   99   aattcagtct~tgagctt~~~~gactgaaatatatt~c~tctcatcttgcata~cgcag~
Dgri   87   ~a~t~cgctatag~ccttcaaaga~t~tattgt~ttacatttcgtatt~~attgc~ccg~

                                      |--DdcΔ2-->
Dmel   115  gttttaaatatca~ctaggt~tctcaa~acta~atttc~~aaa~aataa~~~tcaaatta
Dsim   130  gttttaaatataa~ctaggt~tcacac~acta~agctt~~aaa~aataa~~~taaaatta
Dsec   139  gtttaaaatatca~ctaggt~tcacac~acta~agctt~~aaa~aataa~~~taaaatta
Dyak   115  gcttttagtc~aatc~agaaatatca~~a~~t~attttataat~agt~~~~~ttaatttt
Dere   134  ggttttggtctttgcgatgt~aa~tatta~taca~tgt~~aaacaagaaatttcaa~~~t
Dpse   135  ~~~~~~~~~~~~~~~~~~~~~tctttcaa~~ttttaaatgggc~~~~t~~~~~~~~~~ac
Dper   146  ~~~~~~~~~~~~~~~~~~~~~tctttcaa~~ttttaaatgggc~~~~t~~~~~~~~~~ac
Dana   127  ~~~~~~~~~~~~~~~~~~~~gt~ttt~atggttttaagtttacaatttttttttttttac
Dvir   113  ~~~~~~~~~~~~~~~~~~~~~tat~~tgcactttta~tgcgaaaatgaaacgacgcgtgca
Dmoj   149  ~~~~~~~~~~~~~~~~~~~~gtgtaatgcactttg~tgtgaaaatgaca~g~cgtgtgca
Dgri   136  ~~~~~~~~~~~~~~~~~~~~tttaaatgcactcacat~cg~~aatgacg~t~~~cg~gcc
```

```
    <--DdcΔ1--|
Dmel   165  agttcacagag~ctggcaa~ataaaat~~~~~~~~~~~~gtaatagcttgcatgtatgta
Dsim   180  agcttacagag~ctggcaa~ataaaatatatttttaaatgtaatagcttgcatgtat~~~
Dsec   189  agcttacagag~ctggcaa~ataaaatatatttttaaatgtaatagcttgcatgtatg~~
Dyak   162  aatcaactttg~ctggcaa~ataaaa~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dere   185  atttta~aaatact~ttaatat~agct~~aggcttaaatctt~~~~~~~~~aag~attcc
Dpse   158  ~tatt~~~~~aattaa~~~~~~~~~~ttatt~~~~aata~~~~taa~~~~~~ttgta~~~
Dper   169  ~tatt~~~~~aattaa~~~~~~~~~~ttatt~~~~aata~~~~taa~~~~~~ttgta~~~
Dana   166  atgttttctgatttaagaaatgtattttaatttaaaaaacgagtaaaaaaagttaaaaag
Dvir   150  aaatatcaaacaaca~~~~~ttgc~~~gctt~~~gtcaacaaatacaatttaatatatat
Dmoj   187  aaaaaaaaaaaaaaaattatttttttggttt~~~gttagc~a~~a~aa~gtac~atttaa
Dgri   169  aaatatcaaacaata~~~~attgtgttgtcaacagattga~a~ta~aa~~taa~atttaa

Dmel   211  t~atatatatattttttaaattctaaataaatccatggaaaataaagcctttgatatcc
Dsim   234  ~~~~ttatatattttttt~aattataaataaatccatggaaaataatgccttcgatatcc
Dsec   245  t~atttatatattttttaaaattataagtaaatccatggaaaataatgccttcgatatcc
Dyak   185  ~~~~~tatatattttt~~~~~~~~~~~~~~~~~~~gtggaaa~~~~~~tctccgatatcc
Dere   230  tgagatgta~ataactggaa~~~~~aaataagcc~~~~gaaa~~aatcccgcggatatcc
Dpse   185  ~~~~~~t~taatttattgg~~~~att~~~ttat~tggaaagaaaaa~~~~~~~~~~~~~~
Dper   196  ~~~~~~t~taatttattgg~~~~att~~~ttat~~ggaaagaaaaa~~~~~~~~~~~~~~
Dana   226  aaaaggtataaattattgttaatattattttaaatagcaagagattattataagactata
Dvir   199  atatatataaat~~a~~cacacac~acctgca~agtgt~~ttattt~at~~tacaaatta
Dmoj   238  at~~~tat~~at~~g~~cacac~cga~~ggcacact~t~gttattttatagtacaaatta
Dgri   219  ata~~~aaaaatgcagacacacac~acctgtt~a~t~taattattt~at~~aacaaattt

Dmel   270  agt~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dsim   290  agt~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dsec   304  agt~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dyak   216  agt~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dere   278  agt~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dpse   216  ~~~~~~gtttccataaactt~~ct~taga~~~~~~~~tctgccca~~~~~~~~~~~~~~~
Dper   226  ~~~~~~gtttccataaactt~~ct~taga~~~~~~~~tctgtcca~~~~~~~~~~~~~~~
Dana   286  agtactgtttccttgtagttatctctaaaaactctaatctgtaaggagagattaggatta
Dvir   248  att~~~~~~~~~~~~~~tccacagtt~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dmoj   284  att~~~~~~~~~~~~~~tattcaatt~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dgri   269  aatgcttctacattctacattcagtt~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

Dmel   272  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dsim   292  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dsec   306  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dyak   218  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dere   280  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dpse   244  ~~~~~~~tt~~~~c~~~~~~~~~~~~~~tagtt~~~~~~cgtca~~~~~~~~~~~~~~~~
Dper   254  ~~~~~~~tt~~~~c~~~~~~~~~~~~~~tagtt~~~~~~cgtca~~~~~~~~~~~~~~~~
Dana   346  tatgtgtttgaaacatcctatcactccttaggtaaaatccttctttggtctagcgataag
Dvir   259  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dmoj   295  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dgri   294  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
```

```
Dmel   272  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dsim   292  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dsec   306  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dyak   218  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dere   280  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dpse   258  g~tag~~~~~~~~~~~c~ttccaaga~acataattcaaga~tc~gagctgtcccagctc~
Dper   268  g~tag~~~~~~~~~~~c~ttccaaga~acataattcaaga~tc~gagctgtcccagctc~
Dana   406  gatagaataagattctcatt~ttagctacctaagtcaagactttgaaaaat~ataaatca
Dvir   259  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dmoj   295  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dgri   294  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

Dmel   272  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dsim   292  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dsec   306  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dyak   218  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dere   280  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dpse   301  tgaaccattccat~~~~~~~~~~tcagag~g~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dper   311  tgaaccattcaat~~~~~~~~~~tcagag~g~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dana   464  t~ttcccttcaaaaagaaatagatcagagagccagagattgttaaaattatataaaaaag
Dvir   259  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dmoj   295  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dgri   294  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

Dmel   272  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dsim   292  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dsec   306  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dyak   218  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dere   280  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dpse   320  ~~~~~~~~~~~~~~ggtcttctt~~~~~~~~~~~~~~~~~~~~~~acagctttgactggg
Dper   330  ~~~~~~~~~~~~~~ggtcttctt~~~~~~~~~~~~~~~~~~~~~~acagctttgactggg
Dana   523  tatttgtttttaaaggtcttaattaataaaaagagaagggcataatttcaaatcaatagg
Dvir   259  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dmoj   295  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dgri   294  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

Dmel   272  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dsim   292  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dsec   306  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dyak   218  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dere   280  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dpse   345  cccgagtttaatgtcctgc~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dper   355  cccgagtttaatgtcctgc~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dana   583  taattatacataaaaatattaaacaacatgatgtaacaattatctctctttatttattca
Dvir   259  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dmoj   295  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dgri   294  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
```

```
                                               AFB
                                             -------
                                      <--DdcΔ2--|  |--DdcΔ3-->
Dmel   272  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~tactgattcagcgccca~~attaatgcatg
Dsim   292  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~tactgattcagcgctca~~atcaatgcatg
Dsec   306  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~tactgattcagcgccca~~attaatgcatg
Dyak   218  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~tactgattcagacccca~~atcaatgcatg
Dere   280  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~tactgattcagcccgca~~attaatgcatg
Dpse   363  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ta~ctgagaca~cgcgca~aattaatgcatg
Dper   373  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ta~ctgagtca~cgcgca~aattaatgcatg
Dana   643  gataattataataaaatatattaacaaaagatctgtattag~aagctcaact~ttacttg
Dvir   259  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~tcagttcagcga~~~~~~~~t~~~gagc
Dmoj   295  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~tcagttcagcgt~~~~~~~~tt~tgaga
Dgri   294  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~tcagtttagcgaatgggcattgattaga

Dmel   301  ttccaaaaaagt~~gtcaaaaaac~~~~~~gtgcacaaa~~~~~~~~~~~~~~~~~~~~~
Dsim   321  ttccaaaaaagt~~gtcaaaaa~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dsec   335  ttgcaaaaaagt~~gtcaaaaaat~~~~~~gtgcacaa~~~~~~~~~~~~~~~~~~~~~~
Dyak   247  ttccgaaaaagt~~acaaaataaactctccgtgcagaacgtgc~~~~~~~~~~~~~~~~~
Dere   309  tttcgaaaaagtaaggaacaaa~~ctaaccgtgcagaacgtgc~~~~~~~~~~~~~~~~~
Dpse   391  ~ttccaaaat~~~~~~aaca~accaaaaa~~~at~~~~~~ca~~~~c~~~~~~~c~ag~~
Dper   401  ~ttccaaaat~~~~~~aaca~accaaaaa~~~at~~~~~~ca~~~~c~~~~~~~c~ag~~
Dana   701  ctttaaaaatgttttaaagatataaaaaagatattgtttgccaattcaggttgtcgagtc
Dvir   276  ~~~~cga~~~~~gggtagc~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dmoj   315  tttgcgatt~gagggaagc~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dgri   322  ~ttgcgagtcgagtgtaac~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

Dmel   331  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~tcaaa~~~~~~~~~~~~~~~~~
Dsim   340  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~g~~~~~~~~~~~~~~~~~
Dsec   364  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~a~tcaag~~~~~~~~~~~~~~~~~
Dyak   287  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ataag~tcaac~~~~~~~~~~~~~~~~~
Dere   349  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~acaaa~tcaaa~~~~~~~~~~~~~~~~~
Dpse   421  aagagc~g~~~~~~~~~~c~~~~~~~aa~~taacaaa~tcaaag~~a~gaatcgcaca~a
Dper   431  aagagc~g~~~~~~~~~~c~~~~~~~aa~~taacaaa~tcaaag~~a~gaatcgcaca~a
Dana   761  actctcagattaattgaacttccaaaaaagtaacaaaatcaccgctaagaa~cgcacaca
Dvir   286  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dmoj   332  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dgri   340  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

Dmel   336  ~~~cgagagctgaatttgtttttacgacagcggctgcgattcgaagttcagcggctgcgg
Dsim   341  ~~~cgagtccagaatgtgtttttacgacagcgactgctatttgaagttcagcggctgcgg
Dsec   370  ~~~cgagtgcagaatgtgtttttacgactgcggctgcgattcgaagttcagcggctgcgg
Dyak   297  ~~~cgggtgcgggatgtatt~~~~~~aaaaatacattctcagtccgccgattgagattga
Dere   359  ~~~cgggtggagaatgt~~~~~~~~~~~~~~~~~~~~~~~~~~atttttacataccgccg
Dpse   456  aca~~agttgagaatgttgcagtccggagccacggccgagcggct~~~~~~~~~~~~~~~
Dper   466  aca~~agttgagaatgttgcagtccggagccacggccgagcggct~~~~~~~~~~~~~~~
Dana   820  a~atcag~tgaaaatgatccgaaatccgagcaagtgcagaatttatttttaaaaaacagc
Dvir   286  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dmoj   332  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dgri   340  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
```

```
                                         TTK
                                       --------
                                     GRH
              <--DdcΔ3--|          -------
Dmel   394  act~~~~~gcgat~~~~~~~~tgaaccggtcctgc~~~~~~~~~~~~~~~~~~~~~~~~~
Dsim   399  act~~~~~gagat~~~~~~~~tgaaccggtcctgc~~~~~~~~~~~~~~~~~~~~~~~~~
Dsec   428  act~~~~~gagat~~~~~~~~tgaaccggtcctgc~~~~~~~~~~~~~~~~~~~~~~~~~
Dyak   349  aat~~~~tgagat~~~~~~~~tgaaccggtcctgc~~~~~~~~~~~~~~~~~~~~~~~~~
Dere   391  actgagatgagat~~~~~~~~tgaaccggtcctgc~~~~~~~~~~~~~~~~~~~~~~~~~
Dpse   498  ~~~~~~~~~~~~~~~~~~~~~tgaaccggtcctgcttgggccgtacccgtgcccgatg~~
Dper   508  ~~~~~~~~~~~~~~~~~~~~~tgaaccggtcctgcttgggccgtacccgtgcccgatg~~
Dana   878  caaaaccgaagttcagcggtgcgaaccggtcctgc~~~~~~~~~~~~~~~~~~~~~~~~~
Dvir   286  ~~~~~~~~~~~~~~~~~~agggtgaaccggtcctgcggctgct~~~~~~~~~~~~~~ctcg
Dmoj   332  ~~~~~~~~~~~~~~~~~~gagtgaaccggtcctgcggctgca~~~~~~~~~~~~~~ctcg
Dgri   340  ~~~~~~~~~~~~~~~~~~aggtgaaccggtcctgcggctactctgagagctcggagctcg

Dmel   415  ~~~~~~~~~~ggaattggcagcgctgctggacgg~~~~~~~~~~~~~~~~~~~~~~~~~~
Dsim   420  ~~~~~~~~~~ggaattggcagcgctgctgggcgg~~~~~~~~~~~~~~~~~~~~~~~~~~
Dsec   449  ~~~~~~~~~~ggaataggcagcgctgctgggcgg~~~~~~~~~~~~~~~~~~~~~~~~~~
Dyak   371  ~~~~~~~~~~ggaattggcagcgctgctgggcgg~~~~~~~~~~~~~~~~~~~~~~~~~~
Dere   417  ~~~~~~~~~~ggaattggcagcgctgctgggcgg~~~~~~~~~~~~~~~~~~~~~~~~~~
Dpse   535  ~~~~~~cggtggctgtgggct~gctgctgg~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dper   545  ~~~~~~cggtggctgtgggct~gctgctgg~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dana   912  ~~~~~~~~~~~~~~~~~~~~gaaatgccga~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dvir   315  aagctc~~~~~~~~~~~~gctgctc~g~~cagc~acgcagctccaa~cgaggcgc~~~~~
Dmoj   361  cagctc~~~~~~~~acaaagct~cacag~~agcgaggcggagcgaatcgaggcgag~gcg
Dgri   383  gagctcgcacgactcacagct~cactgctca~ccactcgcagc~tctc~a~tctggtgcg

                TATAAAA
                -------
Dmel   439  ~~gctttaaaagccatggccaagagcgggcagcgc
Dsim   444  ~~gctttaaaagccatggccaagagcgggcagcgc
Dsec   473  ~~gctttaaaagccatggccaagagcgggcagcgc
Dyak   395  ~~gctttaaaagccatggccaagagcgagcagcgc
Dere   441  ~~gctttaaaagccatggccaacagcgagcagcgc
Dpse   558  ~~gctttaaaagccttggccaagtgcgagcagcgc
Dper   568  ~~gctttaaaagccttggccaagtgcgagcagcgc
Dana   922  ~~gctttaaaagcctgggccaagagcgagcagcgc
Dvir   352  ~tgctttaaaagcgccgcccgaagtgcgaccagcactcagttggctaac
Dmoj   409  atgctttaaaaggtcttgccgaacgacgagcagcac~~~~~~~~~~~~~
Dgri   438  atgctttaaaagctttggccgagcgacgagcagcactcagttggctaac
```

pleSubBMin

```
                               CREB
                              --~------
Dmel   1     ggttt~~~~gcagcacgtg~acgtaaaata~~~aactgaaaaa~~~caaac~~~~~~~~~
Dsim   1     ggttt~~~~gccgcacgtg~acgcaaaata~~~aactgaaaaa~~~caaac~~~~~~~~~
Dsec   1     gcttt~~~~gccgcacgtg~acgcaaaata~~~aactgaaaaa~~~caaac~~~~~~~~~
Dere   1     ggttt~~~~gccgcacgtg~acgcaaaaca~~~aactgaaaaa~~~caaactgaaagaaa
Dyak   1     ggttt~~~gctcgcacgtg~acgcaaaata~~~aactgaaaaa~~~caaaccga~~~~~~
Dpse   1     tgttt~~~~gtcgcacgtggacgcaaaatc~~~aagagaaaagg~gaaaatcaaan~~~~
Dana   1     ggttt~~~~ggcgcacgtg~acgtaaaatcga~~~~~~~~~~~~~~caaac~~~~~~~~~
Dvir   1     tgatt~~~~gcggcacgtg~acgtaaaacca~aaacaaaattcacaaagccc~~~~~~~~
DMoj   1     cgggaacggaccgcacggggacgcaaaggcgcaaac~aaaaaaaaa~tcac~~~~~~~~~
Dgri   1     tgagaa~tcac~~cacgtg~acgcaaaatca~aaac~aaaaaatcaccacccaacgatgt

Dmel   40    ~~~~~~~~~~~~~~~~~~agaaaaaacaaacaaaaaatg~~tccaaaacc~~aaaaca~a
Dsim   40    ~~~~~~~~~~~~~~~~~aaaaaactgg~cataaaaca~~~~~~~~~aacc~~aaaaca~a
Dsec   40    ~~~~~~~~~~~~~~~~aaaaaaactgg~cataaaaca~~~~~~~~~~~~~~~~~aaca~a
Dere   50    ctaaaaaaaaaac~aaacaaaaaatgg~cataaaaca~~~~~~~~~aacc~~aaaaca~a
Dyak   44    ~~~~~aaaaaaggcaaacaaaaaatgg~cataaaaca~~~~~~~~~aacc~~aaaaca~a
Dpse   48    ~~~~~nnnngggcaaaatgtag~~tag~~ataaaaaa~~~~~ccataaccagaaaacaaa
Dana   32    ~~~~~aaaaaatcata~~~~~~agtgg~~~~~aataagccaatctaaacc~aa~aaca~a
Dvir   46    ~~~~~~~~~~~~~~~~~aaaagacaaatcaaaaaa~~~~~~~~~~~caaaaaagaaaaga
DMoj   49    ~~~~~~~~~agg~~~~~~~~~~~~~~~~~agtaaag~~~~~~~~~aaacccaaaaact~~
Dgri   54    ~~~~~ccaaaggc~~aaatcaaaata~~ataaa~~~~~~~~~~~~~~~~~cacaaacaaa

Dmel   78    aca~~~~aaat~~~acgaaa~~~~~~~~tgcgaa~~~cgag~~~~~gcgcgcggctattt
Dsim   71    aca~~~~aaat~~~acgaaa~~~~~~~~tgcgaa~~~cgag~~~~~gcgcgcggctattt
Dsec   66    aca~~~~aaat~~~acgaaa~~~~~~~~tgcgaa~~~cgag~~~~~gcgcgcggctattt
Dere   96    aca~~~~aaat~~~acgaaaatacgaaatgcgaa~~~cgag~~~~~gcgcgcggctattt
Dyak   87    aca~~~~aaat~~~acgaaa~~~~~~~~tgcgaa~~~cgag~~~~~gcgcgcggctattt
Dpse   95    cca~~~~aaat~~~gcgaac~~~~gagacgcgaggcgcgaggcgcaccgcgcggctattt
Dana   74    acacgaaaaacaaaacgaaa~~~~~~~~tgcgaa~~~cgagg~~~~~cgcgcggctattt
Dvir   79    aaa~~~~aaaa~~~ac~aaa~~~~~~~~gttgag~~cgcgtt~~~~~cgcggggctattt
DMoj   72    ~~~~~~~~~~caaaacaaa~~~~~~~~~gccagg~cggtgtgc~~~~gtgtttgctattt
Dgri   89    at~~~~~aaa~~~aacaaaa~~~~~~~aaacaga~aagcgt~~~~~gcgcggtgctattt

Dmel   115   tgcatattcaac~~aa~~~~~~~~~~~~~~~~ttttacggctaaatgcgggca~~~~~~~
Dsim   108   tgcatattcaac~~aa~~~~~~~~~~~~~~~~ttttactgctaaatgcgggca~~~~~~~
Dsec   103   tgcatattcaac~~aa~~~~~~~~~~~~~~~~ttttactgctaaatgcgggcc~~~~~~~
Dere   141   agcatattcatttgaa~~~~~~~~~~~~~~~~ttttactgataaatgcgggca~~~~~~~
Dyak   124   tgcatattcaaatgaa~~~~~~~~~~~~ca~~ttttactgataaacgcgggca~~~~~~~
Dpse   144   tgcatattcaaatgaa~~~~~~~~~~~~caatttttattgctaaatgcgagagatacgag
Dana   118   tgcatattcaaatgaa~~~~~~~~~~~~caa~ttttactgctaaatgcgggcgata~~~~
Dvir   116   tgcatattcaaatgaa~~~~~~~~~~atcaatt~~gacggctaaatgcgagac~~~~~~~
DMoj   109   ggcatattcaaatcaaatccaatcaaatcaatttgtat~~ccaaatgcgaaac~~~~~~~
Dgri   128   tgcatattcaaataaa~~~~~~~~~~aacgattt~~gttggttaa~~~~~~~~~~~~~~~
```

```
Dmel   149  ~~~~~~~~~~ataaaa~~~~acacagcatc~~~~~~~~~~~~~~~~~~~~~~~~agttag
Dsim   142  ~~~~~~~~~~ataaaa~~~~acacagcatc~~~~~~~~~~~~~~~~~~~~~~~~agttag
Dsec   137  ~~~~~~~~~~atat~~~~~~aaacagcatc~~~~~~~~~~~~~~~atgtggtagattca~
Dere   177  ~~~~~~~~~~ataaaa~~~~acacagcatc~~~~~~~~~~~~~~~~~~~~~~~~aggcag
Dyak   162  ~~~~~~~~~~ataaaaacacacacagcatc~~~~~~~~~~~~~~~~~~~~~~~~aggcag
Dpse   192  atgtacggagagatacagatacagatacagatacagttgcagataccagcccagctatag
Dana   160  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~aaaag
Dvir   156  ~~~~~~~~~~~~~aaaaca~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~aagctaa
DMoj   159  ~~~~~~~~~~~~~aaaacaaa~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~gctaa
Dgri   160  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

                                          AFB                    GRH
                                       --------                --------
Dmel   172  ccgg~cgtcatctttggccca~aag~tgtgattcagcaccc~~~~~~~~~aaacgagtt~
Dsim   165  ccgg~cgtcatctttggccca~aag~tgtgattcagc~ccc~~~~~~~~~aaacgagtt~
Dsec   165  ~cgg~cgtcatctttggcccagaacatgtgattcagc~ccc~~~~~~~~~aaacgagtt~
Dere   200  ccgg~cgtcatctttggccca~cag~agtgaatcaga~ccc~~~~~~~~~aaacgagtt~
Dyak   189  ccgg~cgtcatctttggcccc~cag~tgtgaatcagt~ccc~~~~~~~~~aaacgagtt~
Dpse   252  ccgg~cgtcatctttggcccc~cag~~gtgaatcagc~cat~~~~~~~~~aaacgagttc
Dana   166  ccgg~cgtcatctttgg~cca~~~~~~gtgaatcagc~cgcg~~~~~~~~aaacgagtt~
Dvir   170  acgg~cgtcatctttgggcca~~~~~~gtgaatcaac~cgc~a~~~~~~~aatccagttc
DMoj   173  atgaacgtcaggctcgggcca~~~~~~gtgaatcagagccc~atgaacgaaatccagttc
Dgri   161  acag~cgtcatccgctgacca~~~~~~gtgaatcagctctcga~~~~~~~aatccagttc

Dmel   219  gat~~~~cttgga~aagcggag~~~~~~~~~~~gagcggag~at~~~~~~~~~~~~~~~~
Dsim   211  gat~~~~cttaga~aagcggag~~~~~~~~~~~gagcggag~at~~~~~~~~~~~~~~~~
Dsec   213  gat~~~~cttaga~aagcggag~~~~~~~~~~agaggggag~at~~~~~~~~~~~~~~~~
Dere   246  gat~~~~cttaga~aagcggag~~~~~~~~~~~gagcggag~at~~~~~~~~~~~~~~~~
Dyak   235  gat~~~~cttaga~aagcggag~~~~~~~~~~~gagcggag~at~~~~~~~~~~~~~~~~
Dpse   298  gatatgacttaga~aagcgcca~~~~tccgccggtgagggggcagcgacggggacggtag
Dana   208  gat~~~~cttaga~aagcgg~cgtcaggtggaga~gtggtgtcaagacccaccccacctc
Dvir   214  gat~~~~~~~~~~~aagctcta~~~~attggtatt~~ggaga~~~~~~~~~~~~~~~~~~
DMoj   226  gat~~~~~~~~~~~aagcacta~~~~attgg~attggagagtgg~~~~~~~~~~~~~~~~
Dgri   207  gat~~~~~~~~gataagcacta~~~~attgatgt~~~~gaga~t~~~~~~~~~~~~~~~~

                                                             AFB/CREB/EXD
                                                               --------
Dmel   245  ~~tcccccgaca~gcgatcttcg~g~~~~~tttcca~~ca~~~~~~gcacgtgattgaca
Dsim   237  ~~tcccccgccg~gcgatcttcg~g~~~~~tttcca~~ca~~~~~~gcacgtgattgata
Dsec   240  ~~tcccccgccacgcgatcttcg~g~~~~~tttcca~~ca~~~~~~gcccgtgattgata
Dere   272  ~~~~ccccgcct~ccgaacttcg~g~~~~~tctcca~~ca~~~~~~gcacgtgattgaca
Dyak   261  ~~~~ccccgcct~ccgaacttcg~g~~~~~tttcca~~ca~~~~~~gcacgtgattgaca
Dpse   353  agacga~cgcgt~~~~~~cgtccag~~~~~tctctaa~~~~gtg~~gcacgtgattgata
Dana   261  ccc~~~~~~~~~~accgcttcccagaccctgatccaatcacgtg~~ccacgtgattgata
Dvir   238  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~gtgga~~~~~~gcatgggattgata
DMoj   253  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~tgggcatatcgaattatgata
Dgri   233  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~agcatttgtgagattgtctgggtttattgata
```

```
Dmel   289  tatcatcgtcaagtggtgg~~~~~~~~~~~~aag~~~~~~~tggactcccctttt~~~~~
Dsim   281  tattatcgccaagtggtgg~~~~~~~~~~~~atg~~~~~~~tggactcccccttt~~~~~
Dsec   285  tattatcgccaagtggtgg~~~~~~~~~~~~aat~~~~~~~tggactcccccttt~~~~~
Dere   314  tatcatcgccaagtga~~~~~~~~~~~~~~~~~~~~~~~~~tggactcccccttt~~~~~
Dyak   303  tatcatcgccaagtggtgaaggtggacgtggatg~~~~~~~tggactccccctttgg~~~
Dpse   395  catcagcgtgttgtgtgga~~~~~~~~~ttggag~~~~~~~tggagtcccccgcg~~~~~
Dana   309  catcatcgcaaattggtgg~~~~~~~~~~~~~~~~~~~~~~~~act~ccccctcggaatc
Dvir   258  tatca~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~aattccccccccattc~~~~
DMoj   275  tgtc~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~catttagcattctgatcgac~~~~
Dgri   266  tgtgagttggtg~~~~~~~~~~~~~~~~~~~ttttccatacaacacaccacgcccc~~~~

                                                              Hox
                                                              ----
Dmel   324  ~~~~~~~~~~~~~~~~~cgacagctcat~cgccgtgga~~~~~~~~~~~ataata~~~ct
Dsim   316  ~~~~~~~~~~~~~~~~~cgacagctcat~cgccgtgga~~~~~~~~~~~ataata~~~ct
Dsec   320  ~~~~~~~~~~~~~~~~~cgacacctcat~cgccgtgga~~~~~~~~~~~ataata~~~ct
Dere   343  ~~~~~~~~~~~~~~~~~cgacagctcat~cg~aatgga~~~~~~~~~~~atagca~~~ct
Dyak   352  ~~~~~~~~~~~~~~~~~cgacagctcat~cgccatgga~~~~~~~~~~~ataaca~~~cc
Dpse   433  ~~~~~~~~atcgcactctgacagctcat~ccacatgga~cactccgctccggttgtggct
Dana   344  tgaacgta~ccgatcccggccagctcaggccccttggc~~~~~cagctc~~~atcgatct
Dvir   278  ~~~~~~~~~~~~atattgacgggctctga~~~~~~~~~~~~~~~~~~~~~~~~~~cagcg
DMoj   298  ~~~~~~~~~~~~tgacggacggactgattgactgattgaa~~~~~~~~~~~~~~~~~~~~
Dgri   302  ~~~~~~~~~~~~~~~~~~~~acccaaccatccagttgga~~~~~~~~~~~~cagttaac~

Dmel   353  agccga~~~~~~~~~~~~~~~~~~~~~~~~~~~cttcgatttgggaaagcaaaacctc~~
Dsim   345  agccga~~~~~~~~~~~~~~~~~~~~~~~~~~~cttcgatttgggaaagcaaaacctc~~
Dsec   349  agtgga~~~~~~~~~~~~~~~~~~~~~~~~~~~cttcgatttgggaaagcaaaacctc~~
Dere   371  agccga~~~~~~~~~~~~~~~~~~~~~~~tcgacttcgatttgggaaagcaaaacgtc~~
Dyak   381  agccga~~~~~~~~~~~~~~~tgcctccaacgacttcgatttgggaaagcaaaacctc~~
Dpse   483  ~~ccgattgtggctcctattgtggctccgat~tggggcatttgggaaagcaaaacctc~~
Dana   395  ggccgg~~~~~~~~~~~~~~~~~~~~~~~~~~~cttcgatttgggaaagcaaaacttc~~
Dvir   301  cacagctgaaactt~~~~~~~~~~~~~~~~~~gtggcgatttgggaaagcaaaacttc~~
DMoj   326  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~tgtatttgcaatttgggaaagcaatacctcg~
Dgri   330  ag~~~~~~~~~~~~~~~~~~~~~~~~~~~~ggactgtgaattgggaaagcaaaacttcgc

                  Hox       Hox                    EXD            Hox
                  ----     ------                  ----         -------
Dmel   383  ~~~aatattatacaataattagcaacaa~~~~~~~~aaatgattgccac~tcgtaattaa
Dsim   375  ~~~aatattatacaataattagcaacaa~~~~~~~~aaatgattgccatgtggtaattaa
Dsec   379  ~~~aatattatacaataattagcaacaa~~~~~~~~aaatgattgccat~tcgtaattaa
Dere   405  ~~~aatattatacaataattagcaagaa~~~~~~~~aaatgattggcat~tcgtaattaa
Dyak   423  ~~~aatattatacaataattagctacaa~~~~~~~~aaatgattgccat~tcgtaattaa
Dpse   538  ~~~aatattatacaataattagcaacaa~~~~~~~~aaacgattgccat~ttgtaattaa
Dana   425  ~~~aatattatacaataattagcagcaa~~~~~~~~aaatgattgccat~ttgtaattaa
Dvir   340  ~~~gatattatacaataattagcaacaa~~~~~~~~aaatgattgccat~ttgtaattaa
DMoj   359  atatatattatacaataattagcaacga~~~~~~~~aaatgattgccaagttgtaattaa
Dgri   362  ttcgatattatacaataattagcaaaaataaaaaacgaatgattgccattttgtaattaa
```

```
                  Hox
           -      ----
Dmel  432  tgcacataattg~~~~~~~~~~~~~~~ccgcgccagattgctgccgtagaatgtagctcg
Dsim  425  tgcacataattg~~~~~~~~~~~~~~~ccgcgccagattgctgccgtagaatgtagctcg
Dsec  428  tgcacataattg~~~~~~~~~~~~~~~ccgcgccagatagctggcgtaaaatg~~~~~~~
Dere  454  tgcacataattg~~~~~~~~~~~~~~~ccgcgccagattgctgccgtagaacgtacatat
Dyak  472  tgcacataattg~~~~~~~~~~~~~~~ccgcgccagattgctgccgtagaatgtagctcg
Dpse  587  tgcacataattg~~~~~~~~~~~~~~~cc~cg~tagattggtgtagtagcagtag~~~~~
Dana  474  tgcacataattg~~~~~~~~~~~~~~~ctgggccagattggtgccgtagaatgtggctcg
Dvir  389  tgcacataattg~~~~~~~~~~~~~~~~ttagccagatcga~~~~ttagaaatagtattt
DMoj  411  tgcacataattgtagaattgttgaattgtgagccagatcga~~~~ttagaacga~~~~~~
Dgri  422  tgcacataattg~~~~~~~~~~~~~~~~tgagccagatcga~~~~ttagaaata~~~~~~

Dmel  477  ctagaatg~~~~~~~~~~~~~~~~~gagtaaaaatacaaaaaaaaaaaaaaaaaacaaaa
Dsim  470  ctagaatg~~~~~~~~~~~~~~~~~gagtaaaaatac~~~~~~aacaacaaaaaa~taaa
Dsec  465  ~~~~~~~~~~~~~~~~~~~~~~~~~gagtaaaaatac~~~~~~~~~gaaaaaaaa~taaa
Dere  499  tcattatcgcttagattgtcgggta~~~~~~~~cctaccct~gtacggcaaaaaacaaaa
Dyak  517  ctagagtggagtaa~~~~~~~~~~~~~~~~~~~ctaaca~~~aatgaacaaaaa~~~~~a
Dpse  624  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~aatggaaaaaaaa~~~~~
Dana  519  at~~~~~~~~~~~~~~~~~~~~~~~~~~agaatggaaaaaaataaaaataaaaa~~~tac
Dvir  429  gtagtagt~~~~~~~~~~~~~~~~~~~agtagtagcagctgtaatgggaaaaaa~~~~~~
DMoj  460  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ctcgctatgtaatggcaaaaaa~~~~~~
Dgri  455  ~~~~~~~~~~~~~~~~~~~~~~~~~~aaaatactagttaagtaatgggaaaaaa~~~~~~

                                                        Hox        Hox
                                                        ----       ----
Dmel  520  taaaaacag~~~~~~~~~~aattgagaatttttgaaatggactcttaattgtatttatta
Dsim  507  taaaaacag~~~~~~~~~~aattgagaattttcgaaatggactcttaattgtatttatta
Dsec  492  taaaaacag~~~~~~~~~~aattgataattttcgaaatggactcttaattgtatttatta
Dere  550  aaaaaacag~~~~~~~~~~aattgagaatttttgaaatggactcttaattgtatttatta
Dyak  550  aaaaaacag~~~~~~~~~~aattgagaatttttgaaatggactcttaattgtatttatta
Dpse  637  ~~~~~~~~~~~~~~ttaaaaattgagaatttttgaaatggattcgtaattgta~ttatta
Dana  550  aaaaagcagcaaaattaaaaattgagaatttttgaaatggattcgtaattgta~ttatta
Dvir  463  ~~~~~~~~~tgtttataaaaatg~~~~~tttttcaaatggatttttaattgta~ttatta
DMoj  482  ~~~~~~~~~~~~ttgtaaaaatg~~~~ttttttaaaatcgttttttaattgta~ttatta
Dgri  483  ~~~~~~~~~~~~ttataaaaatg~~~~~~ttttaaaatggatttttaattgta~ttatta

                    Hox    Hox     Hox
           -        ----   ----    ----
Dmel  570  attgcgtttaattgaattatgaataatatttatttgc~~~~~~~~~~~~~~~~~~~~~~~
Dsim  557  attgcgcttaattgaattatgaataatatttatttgc~~~~~~~~~~~~~~~~~~~~~~~
Dsec  542  attgcgcttaattgaattatgaataatatttatttgcc~~~~~~~~~~~~~~~~~~~~~~
Dere  600  attgcgcttaattgaattatgaataatatttacctgc~~~~~~~~~~~~~~~~~~~~~~~
Dyak  600  attgcgcttaattgaattatgaataatatttatttgctg~~~~~~~~~~~~~~~~~~~~~
Dpse  683  attgcacttaattgaattatgaataatatttaatttgattcat~~~~~~~~~~~~~~~~~
Dana  609  attgcgcttaattgaattatgaataatattta~gtggccgcct~~~~~~~~~~~~~~~~~
Dvir  509  attgcaccttattcaattatgaatttatatttgatt~~~~~~~~catacaaagcattgcca
DMoj  526  attgcaccttattcaattatgaataatatttgatt~~~~~~~~catacaaaagact~~~~
Dgri  525  attgcaccttattcaattatgaataatattctatttc~~gattcatacaaagcattgcca
```

```
Dmel   606  ~ccccct~~~~~~~~aggaagt~~~~~~~~~~~gccg~~~~~gtaca~~~~~~~~~~~~~
Dsim   593  ~ccccct~~~~~~~~aggaagt~~~~~~~~~~~gccg~~~~ggtaca~~~~~~~~~~~~~
Dsec   579  ~ccccct~~~~~~~~aggaagt~~~~~~~~~~~gccg~~~~tctaca~~~~~~~~~~~~~
Dere   636  ~ccccc~~~~~~~~~aggaagt~~~~~~~~~~~gccg~~~~~gtaca~~~~~~~~~~~~~
Dyak   638  ~ccccct~~~~~~~~aggaagt~~~~~~~~~~~gcag~~tcagtaca~~~~~~~~~~~~~
Dpse   725  ~~~~~~~~~~~~~~~aggaagt~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ggagtc~
Dana   650  ~~~~~~~~~~~~~~~aggaagt~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~gcggtc~
Dvir   561  cagcccttgcagctca~gcatattaaagactacgcaaagaaaataaaaaaaaaaagacag
DMoj   573  ~~~~~~~~~~~~~~~atggttt~~~~~~~~~~~acaa~~aacgaa~~~~~~~~~~~acag
Dgri   583  tagcccttgca~~~~~~~~~~~~~~~~~~~~~~~~~~gcttgcaacaaaagacgacgaaa

Dmel   628  ~~~~~~~~~~gtcctaccacagatccg~~~tattctcggaagtcccc~~~~~~~~~~~ag
Dsim   616  ~~~~~~~~~~gtcctaccacagattcg~~~tattctcggaagtcccc~~~~~~~~~~~ag
Dsec   602  ~~~~~~~~~~gtcctaccacagattcg~~~tattctcggaagtcccc~~~~~~~~~~gac
Dere   657  ~~~~~~~~~~gtcctaccacagattcg~~~tattctcggaagtcccc~~~~~~~~~~~ag
Dyak   663  ~~~~~~~~~~gtcctaccgcagattcg~~~tattctcggaagtcccc~~~~~~~~~~~ag
Dpse   738  ~~~~~~~~~~~~cctgcaacagattcactatattctcggaagtttt~~~~~~~~~~~~tg
Dana   663  ~~~~~~~~~~~~ctagc~acagatttg~~~agttctcggaagtttttcc~~~~~~~~~aa
Dvir   620  aatgg~~~~~~~~~~~~aacagat~~~~~~tattctaggaagttgcca~~~~~~~~~gaa
DMoj   595  aaaaaa~~~~~~~~~cccaaagat~~~~~~tattataggaagttaa~~~~~~~~~~~~aa
Dgri   617  aaaagcacgcagaatagaaaagat~~~~~~tattctaggaagttgcttagtaacacagaa

Dmel   665  caagggcatcc~~~~~~~gacgagg~~~~~~~~~~~~~~~~~~~~gctgg
Dsim   653  caagggcatcc~~~~~~~gacgagg~~~~~~~~~~~~~~~~~~~~gctgt
Dsec   640  caagggcatcc~~~~~~~gacgagg~~~~~~~~~~~~~~~~~~~~gctgg
Dere   694  caagggcatcc~~~~~~~aacgagg~~~~~~~~~~~~~~~~~~~~gctgg
Dyak   700  caagggcatcc~~~~~~~aacgagg~~~~~~~~~~~~~~~~~~~~gctgg
Dpse   775  caagggcacac~~~~~~~gacgagg~~~~~~~~~~~~~~~~~~~~gctgg
Dana   699  caagggcgttc~~~~~~~gacgagg~~~~~~~~~~~~~~~~~~~~gctgg
Dvir   653  caagggcaaaactgtgtacttgagcagccagagagagctc~~gagacggc
DMoj   628  caagggcaaaa~~~~~~actagaggaact~~tgagaactcgactcaaggc
Dgri   671  caagggcaac~~aatagacttgaggaact~~ttagaactcg~~~~aaggc
```

kkv1

```
Dmel   1    caacaaaggattgaagggaaatatatgc~~~~~~~tctgc~~gaatgaattaa~~ata~t
Dsim   1    cagcaacggattggcgggaaatacttggccttagatcgac~~gaatgaattat~~tta~t
Dyak   1    tttatatggtcggaaacgcttccttctgcctgttacatacttttcaacgaatctagtata
Dere   1    ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dpse   1    ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dper   1    ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dana   1    ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dvir   1    CAATTTCCCTTTAGAAATGCCATTGGTGGGGGGAAGTTTTTTTGGGACCCGCCCCGCTTA
Dmoj   1    ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dgri   1    tcaataagcttttatttcgttggatttaaactattaggaatttgaatcatctctcgtatc

Dmel   49   tcactttttaggagatac~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dsim   56   tcactttttggaaggtacatttgttaaaatgt~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dyak   61   ccctttactctacgagtaacgggtataataatcttcccctttgatttaaaggtatgtaa
Dere   1    ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dpse   1    ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dper   1    ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dana   1    ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dvir   61   AATAAATAAAAAATTTTTTTTTTTATAAATTTCATTTTCCCTTGTGTCTGGGGGCGGTAG
Dmoj   1    ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ct
Dgri   61   accttgtgaagtaggcgttagtgaaatgaaaaagcaaagtgatccaatattttggtgcaa

Dmel   66   ~~~~~~~g~~~~~taaactaatttttctgtgcacatataaggctatgtatgcgtttaaat
Dsim   87   ~~~~~~~a~~~~~taaacgaatctttctgtgcacagatacggctgtgtatccgttcgaat
Dyak   121  acaatatattgcataaactgatttttctgtgcacccataaaggtaagtatccgtccgaat
Dere   1    ~~~~~~~~~~~~~taaactgatttttctgtgcaccgatgaaggtatgtatccgcccgaat
Dpse   1    ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~aat
Dper   1    ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~aat
Dana   1    ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~aat
Dvir   121  TGATCAATTATCCAAGTGAAGGGTAATGGAATGCCTAATTCTTATGGCTCTGAACACAAT
Dmoj   3    gattcattatgactcgcaacataaaattcatttacaaatttcagcaatatatttctgcat
Dgri   121  tatatttccacacgcaaaatacatatatacaaactacatacatacatatgtggcttaaaa

Dmel   115  gttttggctttgcggctcagcttagtcagaag~~~~~~~~~~~~ccaccatagaaggcga
Dsim   136  gttttggtcttgctgctcagcttagtcagaag~~~~~~~~~~~~ccgccaaagaaggcga
Dyak   181  attttggttttgcgctccagcataatcagaag~~~~~~~~~~~~cc~~~aaggaaggcaa
Dere   48   gttttggttttgctggccagcttagtcagaag~~~~~~~~~~~~ccgccaaagaaggcga
Dpse   4    gttttggggttttttttcttgtgtttggcttcaagactcgttacccgccaaagaaggcaa
Dper   4    gtttttgggttttttttcttgtgtttggttccaagactcgttacccgccaaagaaggcaa
Dana   4    g~~tttaggttttgaggc~taagtctg~~t~~~~~~~t~g~~~~ccgccaaagaaggtaa
Dvir   181  ATTCATAATAAACATTTGAGCAATATATATGTATTTCAAACTGTTGAGTTGTGCATTGCG
Dmoj   63   gcaaagtacatagagtaaatacatacacaaatacaaatatactgtagtatatatgtcatg
Dgri   181  tatgatttgtgccccaagaaggcgagcgataaaattaaaagccactcaagcaagcacacg

Dmel   163  acaattaaaagctgcgctgcgcggaaaggtagccagaaagacagaagcaaaaaaaagtga
Dsim   184  acaattaaaagctgcgctgcgcggaaaggcagccagaaagacagaagcagaaaaaagtga
Dyak   226  acaattaaaacttgcgctgcgcggactgccagccag~~acac~~~a~ccgaaaaaagtga
Dere   96   acaattaaaacctac~~~~~~~~~~~~~~cagccaggcacac~~~~~~~ataaaaagtga
Dpse   64   acaattaaaacccgcgctccaccaacaggcagc~aagaaat~~~~~~~~~~aaaaagtga
Dper   64   acaattaaaacccgcgctccaccaacagacagc~aagaaat~~~~~~~~~~aaaaagtga
Dana   47   acaattaaaagctgcgctaaaccaactgccagc~cagaaaaaaacacaga~aaaaagtga
Dvir   241  ATTCAAAGCAACGACTTAAATGCCGCACAGTCCAGCAAATACAATGGCAAGAAAAAGTGA
Dmoj   123  cccaagaaggcaaatcaaatggcttacagcccaacacaaaacgaatgcatgaaaaagtga
Dgri   241  acgtgtgtatgctt~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~gaaaaagtga
```

```
Dmel   223   aagtc~agatacttgt~~~~~~~aaaaatgttatagaaataagcgaac~~~~~~~acaat
Dsim   244   aagtc~agatacttgt~~~~~~~aaaaatgttatagaaataagcgaac~~~~~~~acaat
Dyak   280   aagtc~agaaacttgt~~~~~~~aaaaatgttatagtaataagcgaacc~~~~~~a~agt
Dere   135   aagtc~agatacttgt~~~~~~~aaaaatgttatagaaataagcgaac~~~~~~~acagt
Dpse   113   aagtccagatacaagtac~~~aaataaatgttatagaaaaaaaa~cagattgtatacaat
Dper   113   aagtccagatacaagtac~~~aaataaatgttatagaaaaaaaaacagattgtatacaat
Dana   105   aagt~cagatacttgta~~~~aaatgaatgttatagaaataaac~~~~~~~gaacacaat
Dvir   301   AAGTCTA~~~~~~~~~~~~~~~~AAAAATGTTATAATAAAA~~~~~~~~~~~~~~ACAAT
Dmoj   183   aagtctaaagcaaaaacaacaacaatattgttataataaaa~~~~~~~~~~~~~~acgat
Dgri   265   aagtct~~~~~~~~~~~~~~~~aaaaaatgttataataaaa~~~~~~~~~~~~~~acaat

Dmel   268   ctatgcgatactggtg~ctcagataccggagtgtttcatatacacaagcccga~~~~~~~
Dsim   289   ctatgcgatactggtg~ctcagataccggagtggttcagatacacaagcccga~~~~~~~
Dyak   325   ctatgcgataccgggg~ctcagataccgaagtggttcagat~~acaaactcga~~~~~~~
Dere   180   ctatgcgatactgggg~ctcagataccgatgtggttcagat~~acaaactcga~~~~~~~
Dpse   169   ctatgagatacttggc~cacagatacagatacagctgcagctaca~ggccaca~~~~~~~
Dper   170   ctatgagatacttggc~cacagatacagatacagctgcagctaca~ggccaca~~~~~~~
Dana   153   ctatgcgatac~tggctcacagataccgacacagatcca~caagt~tg~~~~~~~~~~~~
Dvir   331   CTATGAATAAGATACTACAAGCCGAGACCCAACAGCAACAGCATTGCCGTCATGC~~~~~
Dmoj   229   ctatgaataagatacttgtagctacaa~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dgri   295   ctatgaataagacaaaacagcagcaaatgcgggaacacagcagcagcagcagcacagcaa

Dmel   319   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~gttg~ctg~ctgcttgga~~~~~~~~
Dsim   340   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~gttgctg~ctg~ctt~~~gga~~~~~~~~
Dyak   374   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~gttgttg~ctg~ctt~~~gaa~~~~~~~~
Dere   229   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~gttg~ctg~ctt~~~gga~~~~~~~~
Dpse   219   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~gcca~ctg~ctcactgggctgcaac
Dper   220   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~gcca~ctg~ctcactgggctgcaac
Dana   197   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ggg~ctg~caaccgttgttgctact
Dvir   385   ~~~CAACAACAACTCAGTGGTTGCAAC~~~T~~~GTTG~CTGGCTGCTG~~~~~~GTTG~
Dmoj   255   ~~~cagcaaccgctcggtagttgcaac~~~t~~~attg~ctggctgttgaaaattgttg~
Dgri   355   cagcaacaacaactcagtagttgcaacagttgttgttggctggtg~~~~~~cctg~~~~~

Dmel   335   ~~~~~gtggcgcacaaattgattgctacacttaagccatctagtcaaaacgggaacgcgg
Dsim   356   ~~~~~gtggcgcacaaattgattgctacacttaggccaactagtcaaaacgggaacgcgg
Dyak   390   ~~~~~atggcgcacaaattgattgctacacttaagccatctagttaaaacggc~~~~~~~
Dere   242   ~~~~~gtggcgcacaaattgattgctacacttaagccaactagtcaaaacgggaacgcgg
Dpse   244   t~~~ggtggcgcataaattgattgctacatttaagccaactagtcaaaacgggaacacgg
Dper   245   t~~~ggtggcgcataaattgattgctacatttaagccaactagtcaaaacgggaacacgg
Dana   221   tggggctggcgcataaattgattgctacatttaagccaactagtcaaaacgggaacacgg
Dvir   428   ~~~~~~~~~~~~~~AAATTGATTGCTACATTTA~~~~~~~~~~~~ACAACGGAAACAA~~
Dmoj   304   ~~~~~~~~~~~~~~aaattgattgctacatttta~~~~~~~~~~~~acaacggaaacga~~
Dgri   403   ~~~~~~~~~~~~~~aaattgattgctacatttta~~~~~~~~~~~~acaacggtaacga~~

                                                                        --
Dmel   391   c~~~~~aaccga~~~~~~~~~agc~~~~~~~~~~~~~~~~~~~~~~~~~ggccacgccaa
Dsim   412   c~~~~~aaccga~~~~~~~~~agc~~~~~~~~~~~~~~~~~~~~~~~~~ggccacgccaa
Dyak   438   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ggccccgccaa
Dere   298   c~~~~~agtcgc~~~~~~~~~agc~~~~~~~~~~~~~~~~~~~~~~~~~ggccacgccaa
Dpse   301   c~~~~~cgaaccggcccagacctg~~~~~~~~~~~~~~~~~~~~~~~~~ggccacgccag
Dper   302   c~~~~~cgaaccggcccaggcctg~~~~~~~~~~~~~~~~~~~~~~~~~ggccacgccag
Dana   280   ~~~~~~c~aatcg~~~~~aacc~~~~~~~~~~~~~~~~~~~~~~~~~~~ggccacgccag
Dvir   460   ~AATCGAACCAATCGAACCAAACGTGT~~~~~~~~~~~~~~~CCAAACCGGCCACGCCAA
Dmoj   336   ~aaacgaaccaatcgaaccgacgaatggaccgacccacccagccaaaccggccacgccaa
Dgri   435   ~aaccgcacaaatcgaaccaatacg~~~~~~~~~~~~~~~~~~caaaccggccacgccaa

              GRH
```

```
                  ------
Dmel    412   ccagctgcctactgc~~~~~~~~tattgccgatcgccgatcgctggttgctagttgctga
Dsim    433   ccagctgcctactgcctgctgctattgccgatcgccgatcactg~~~~~~~gttgctgat
Dyak    450   ccagctgcctactac~~~~~~~tcttgccgatcgccgatt~~~~~~~~~~~~~~~~~~~~
Dere    319   ccagctgcctactgc~~~~~~~tattgccgatcgacgatt~~~~~~~~~~~~~~~~~~~~
Dpse    331   gcagcagctgccggactgctgccggactgctgct~~~~~~~~~~~~~~~~~~~~~~~~~~
Dper    332   gcagcagctgccggactgctgccggactgctgct~~~~~~~~~~~~~~~~~~~~~~~~~~
Dana    302   ccagctgctgattgc~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dvir    505   GCTGTCCGGCTGTCCGGCAGTCCGACAACTGAGTCCAACTTGTATTATAAGCGTTGCTAA
Dmoj    396   gccggacagccagcaacagcaacagcagctgccgcagcaactgcgtccgacttgtatcat
Dgri    477   gctcttggccaacaacaaaactgcagcggccaagactcgtatcatgag~~~~~~~~~~~~

Dmel    464   ttgctgattgccgatggccggacaaccagtttt~~~~~~~~~~~~~~~~cggcggacaac
Dsim    486   t~~~~~~~~~~~~~~~gccggacaaccagtttt~~~~~~~~~~~~~~~~cggcggacaac
Dyak    482   ~~~~~~~~~~~~~~~~gccggacaaccagtttt~~~~~~~~~~~~~~~~cggcggacaac
Dere    351   ~~~~~~~~~~~~~~~~gccggacaaccagtttt~~~~~~~~~~~~~~~~cggcggacaac
Dpse    364   ~~~~~~~~~~~~~~~~~~cggacaaccagttct~~~~~~~~~~~~~~~~cggtggccaac
Dper    365   ~~~~~~~~~~~~~~~~~~cggacaaccagttct~~~~~~~~~~~~~~~~cggtggccaac
Dana    316   ~~~~~~~~~~~~~~~~~~cggacaaccagtttt~~~~~~~~~~~~~~~~tggtggacagc
Dvir    565   C~~~~~~~~~~~~~~~~~~~~~CAACCAGTTTTTAATA~~~~~~~~~~~~~~~~GAGTAC
Dmoj    456   aagacag~~~~~~~~~~~~~~~agaccagtttc~~~~~~~~~~~~~~~~tgctggagtac
Dgri    524   ~~~~~~~~~~~~~~~~~~~~~~caaccagtttaacatttttttttttttaatcgagtac

Dmel    508   aggaaagccagcgaactgcg~~~~~~gccaagaaatatgcgccaatat~~~~gactgaaa
Dsim    515   aggaaagccagcgagctgcg~~~~~~gccaagaaatatgcgccaatat----gagtgaaa
Dyak    511   aggaaagccagcgaggtacg~~~~~~gccaagaaatatgcgccaatat----gactgaaa
Dere    380   aggaaagccagcgagctgcg~~~~~~gccaagaaatatgcgccaatat----gactgaaa
Dpse    391   aggaaaagcagccgcctcttggccaagccaagaaatatgcgccaataa----gagtaaaa
Dper    392   aggaaaagcagccacctcttggccaagccaagaaatatgcgccaataa----gagtaaaa
Dana    343   aggaaaa~~~~~~~~~~~~~~~~~~~~ccaagaaatatgcgccaataaataagagtaaaa
Dvir    588   CACAGGAACAGTGAGCGACTTGAAACGTA~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dmoj    485   aacaggaacaataagcgactttaaaag~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dgri    563   aggaaaatgc~~gatcgacttgaaaag~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

Dmel    558   tagtagccatta~~~~tgcgaaaa~~attgcat~gcgaagcaa~~gccgg~aacgaacgg
Dsim    565   tagcagccatta~~~~tgcgaaaa~~attgcag~gcgaaccaa~~gccgg~aacgaacgg
Dyak    561   aagcaggcatta~~~~tgagaaaa~~attgcag~gcgaaccaa~~gccgg~aacgaacgg
Dere    430   tagcaggcatta~~~~tgcgcaaa~~attgcag~gcgaaccaa~~gccgg~aacgaacgg
Dpse    447   ttgt~~~~~~~~~~~~~~~~~~aa~~attgcaa~tcgaagcaa~~gccgg~aacgg~~~~
Dper    448   ttgt~~~~~~~~~~~~~~~~~~aa~~attgcaa~tcgaagcaa~~gccgg~aacgg~~~~
Dana    383   ttgt~~~~~~~~~~~~~~~~~~aagcatt~~aagt~gaa~aaattgcaggcaaagttaag
Dvir    616   ~~~~~~~~~~~~~~~~TGCGAAAAACATGCCCAAACATAAAA~~~~GATAAAAGT~AGCA
Dmoj    511   ~~~~~~~~~~tggcattgcgcaaaacatgcccaaa~caatgactgcgataatggccagca
Dgri    587   ~~~~~~~~~~ta~~~~tgcgcaaaacatgcccaaacaatgacaaataatggcagcagtaa

Dmel    608   c~~gggtgaattgg~~~tgg~~tgacatggacaagacggggcaagttttggagtggcgat
Dsim    615   c~~gggtgaattgg~~~tgg~~tgacatggccaagacggggcaagttttggagtggcgat
Dyak    611   c~~gagtgaa~t~~~~~tgg~~tgacatggccaagtcgggacaagttttggagtggagat
Dere    480   c~~gagtgaa~t~~~~~tgg~~tgacatggccaagacggggcaagttttggggtggggat
Dpse    478   ~~~~aacgaagcgaca~tgg~~tgacatggctgagatggggaaggcaggctctg~at~~~
Dper    479   ~~~~aacgaagcgaca~tgg~~tgacatggctgagatggggaaggcaggctctg~at~~~
Dana    421   ccggaacgaa~ctacaatgg~~tgacatagccaagatg~~~~tggcaggctctggaggag
Dvir    656   ACAAATGCTGTCAGAATAGTGACATAGGAAGGTAATATGGACTAAACCGGATTATACAAT
Dmoj    561   atgaatatgatcagaacggtgaacatggcagggcagcatgatctagctgcgatcaatgtg
Dgri    634   atgtggccagaatgccgccatagaaaggcaatatgaaccgaacgggtattaaatggctta
```

```
Dmel   661   c~~~~~~~~~~~~~~~~~~~~~~cggggagggacaaccatgggtgtgtgggttcaatgcg
Dsim   668   c~~~~~~~~~~~~~~~~~~~~~~cggggagggacaaccatgggtgtgtgggttcaatgcg
Dyak   661   c~~~~~~~~~~~~~~~~~~~~~~cggggagggacatccatgggtgtg~~ggttcaaagcg
Dere   530   c~~~~~~~~~~~~~~~~~~~~~~cgggaaggggcatccatcggtgtg~~ggttcgaggcg
Dpse   528   tcggtggcacggtagcaaggggtcgggcaggggtt~ccatgggtgtgggttcagatc~ac
Dper   529   tcggtggcacggta~~~~~~~~~~~~gcaggggtt~ccatgggtgtgggttcagatc~ac
Dana   474   t~ggagggac~~t~ggaatgggt~~~g~tggag~~~ccagtggt~~~~~~~cag~tccac
Dvir   716   TGTTTTAAAGTAATTCTGAACTAAATGGAGAATAGTTCAGGTATATCGGTTTCAGGTATT
Dmoj   621   ggttcaatcaattgtgctataaatgctgtatatatttacttgtaaaatggttc~~~~~~~
Dgri   694   accactgcaatcgaggtgtgggttcaatgggcattgactgactgaccactggt~~~~~~~

Dmel   699   cagtggtcagctggctggggttggctcaagtcaacttactcaaaccatcgatttatgctt
Dsim   706   cagtggtcagctggctgggattggctcaagtcgagttactcaagccatcgatttatgctt
Dyak   697   tagtggttagcttgctgggattggct~~~~attgattccttaagccattgatttctgctt
Dere   566   cagtggtcagctggctgggattggct~~~~attggttacttaagccactgatttatgctt
Dpse   586   tggaattgaccatttaagtt~~~~~~~~~~~~~~~~~~~~~~~~~~attgatttatgcat
Dper   575   tggaattgaccatttaagtt~~~~~~~~~~~~~~~~~~~~~~~~~~attgatttatgcat
Dana   515   tgg~atcggccatttaagct~~~~~~~~~~~~~~~~~~~~~~~~~~attgatttat~~~~
Dvir   776   TGCGCTTAACATATGGTTTTTACATATAAGCAGAGCTAGTGCAATTCTTGAATTCCCCAT
Dmoj   673   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dgri   746   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

Dmel   759   gccacgac~~~aga~gagccacaattgcttgaaggttgcctgcaatccgattgcaggaat
Dsim   766   gccacgac~~~gga~gagccacaattgcttgaaggttgcctgcaatccgattgcaggaat
Dyak   753   gccacaac~~~gga~gagccataattgcttgaaggttacctgcaatccgattgcaggaat
Dere   622   gcctcaac~~~gga~gagccacaattgcttgaaggttgcctgcaatccgattgcaggaat
Dpse   620   gctacggccagagtggccgcgctgctgcctgaaggttgttcgcgatctaa~~~caggcag
Dper   609   gctacggccagagtggctgcgctgctgcctgaaggttgttcgcgatctaa~~~caggcag
Dana   543   ~~~~~~~~~~~~gt~~~~~~~~t~~tatttgaaggttgcc~~~~~t~~~~~~~~~~~~~g
Dvir   836   GCAGTTCAACAAACCTCAATTTCCTTTCAAGACCTGTCAATTAAGAGTTTTTGC~~~~~~
Dmoj   673   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dgri   746   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

               GRH                                     AFB
             -------                                 -------
Dmel   815   accagttgctatcgatccggg~~~~~~~~~~~~~~~~~~~tgattcatggctgcgcatgc
Dsim   822   accagttgctatcgacccggg~~~~~~~~~~~~~~~~~~~tgattcatggctgcggatgc
Dyak   809   accagctgctatcgacccgga~~~~~~~~~~~~~~~~~~~tgattcatggctacggatac
Dere   678   accagttgctatcgccccgga~~~~~~~~~~~~~~~~~~~tgattcatggttgctgatgc
Dpse   677   gaatgccaccttgaggggatt~~~~~~~~~~~ggcttgtatgaatca~~~~~~~~~tagg
Dper   666   gaatgccaccttgaggggatt~~~~~~~~~~~ggcttgtatgaatca~~~~~~~~~tagg
Dana   564   gaat~~~~~ct~~~~~~gattgccggagtactagtacagatgactca~~~~~~~~~~~~~
Dvir   889   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dmoj   673   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dgri   746   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

Dmel   856   cagacgatccatcttggaggtgcacctagcagctagcacccagctcgaa~agatcgccaa
Dsim   863   cagacgatccatcttggagctgcacctagcagctagcacccagctcgaa~agatcgccaa
Dyak   850   cagacaatccatcttggagccgc~~~~~gca~~tagcacccagttcgaa~agatcgctta
Dere   719   cagacgatt~~~~~~~~~~~~~~~~~~~~~~~~~~~cacccaactc~aagagatcgccta
Dpse   717   tatagatctttggccatacgagtattcctagaaattaagttcaattcagttaaagaaaag
Dper   706   tatagatctttggccatacgagtattcctagaaattaacctaagttcagttaaagcaaag
Dana   599   ~~~~~~~~caggcatttga~~~~tgtgtggaagctgcacccaacttgagttaaag~~~~~
Dvir   889   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dmoj   673   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dgri   746   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
```

```
Dmel   915  tcga~gct~tggga~gagcttactccgtcgatttgagcactggataatttgtaaagtttt
Dsim   922  tcga~gct~tggga~gagcttactccgcagatttgagcactggataatttgtaaagtttt
Dyak   902  tgaa~gctatcaaaagagcttac~actctgctttaa~~~~~~~~~~~~~~gcaaagtttt
Dere   750  ~ccatgct~ttgga~gagcttactccgctgatttaagctctggataatttgtaaagtttt
Dpse   777  catttgctaaagtcgaggcagat~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dper   766  catttgctaaactcgaggcagtttatggggataaagagattctaaaagaagctatatatt
Dana   642  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dvir   889  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dmoj   673  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dgri   746  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

Dmel   972  ggttatcacattttcattgtttaacaatttgcac~~~~~~~~~~~~~~~~~~~~~~~~~~
Dsim   979  ggttatcacattttcattgtttaacaatttgcaa~~~~~~~~~~~~~~~~~~~~~~~~~~
Dyak   946  agttatcacattttc~ttgcctaacaatttgaaa~~~~~~~~~~~~~~~~~~~~~~~~~~
Dere   808  agttatcacattttcattgtttaacaatttgcaa~~~~~~~~~~~~~~~~~~~~~~~~~~
Dpse   799  ~~~~~~~~~~~~~~~~~~~ctacttgaccctacaaacaggtcagtatctatggaagtatc
Dper   826  ctgggaattataaataaaactacttgactctacaaacaggccagtatctaaggaagtatc
Dana   642  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~taagatgatattcaagt
Dvir   889  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dmoj   673  ~~~~~~~~~~~~~~~~~~~~~~~~~~gagaatgt~~~~~~~~~~~~~~~~~~~~~~~~~~
Dgri   746  ~~~~~~~~~~~~~~~~~~~~~~~~~~aacgggtt~~~~~~~~~~~~~~~~~~~~~~~~~~

Dmel   1005 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ttaacattgagccgataata
Dsim   1012 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ttaacattgagccgataata
Dyak   978  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ttaacataca~ccgataata
Dere   841  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ttaacataaaaccgataata
Dpse   841  tttatggtatctatggctatcatctgtagagtagtttcaattaacattcgactgataatt
Dper   886  tttgtggtatctatggctatcatctgtagagaagtttcaattaacattcgactgataatt
Dana   660  ctgtttcaataaactttgagtggatagacaacagtcttaattaacagtgagttgataatt
Dvir   889  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dmoj   681  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~attatgtttgagaatggttt
Dgri   754  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ggcaataggaagtattgccg

Dmel   1026 aatgctaagtttg~aatg~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dsim   1033 aatgctaagtttg~attg~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~aa
Dyak   998  ggtggtgagttag~agtgtctggactgaatctttaaaatacttcttttaaa~~~~~~~~~
Dere   862  tatgctaag~ttgcagtg~~~~~~~~~tatatataaaacaccactttt~aa~~~~~~~~~
Dpse   901  tacatttaaagctaatgaaacaaatatacaaaagcaaattcgaatttccagcttttcaaa
Dper   946  tacatttaaagctaatgaaacaaatatacaaaagcaaattcgaagttccagcttttcaaa
Dana   720  ttcctt~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dvir   889  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dmoj   702  gagaacgattgctgcatactcgagaagtattacc~~~~~~~~~~~~~~~~~~~~~~~~~~
Dgri   775  acgagctataaaccaaaatggacgagtggtgggacccggaatgggaagttttttaaaatg

Dmel   1043 ataaaacattcgaaactattg~aaaat~~ttgaag~~~~~~~~tttc~aaa~tatttttc
Dsim   1052 aaaacacattcgaaactattg~aaagt~~ttcaag~~~~~~~~tttc~aaa~tattttcc
Dyak   1047 ~~~~~~~~~~~~aaaccttttacaaaaacttaa~~~~~~~~~~aaac~aaaatattttcc
Dere   901  ~~~~~~~~~~~~aaaccctttaccaaaacttaa~~~~~~~~~~aaac~aaaatacttcac
Dpse   961  ccagtaaaaggaaaagcaaaaactttctgagata~~~~~~~~~~~~~aacaaaactaaag
Dper   1006 ccagtaaagggaaaagcaaaaactttctgtgatatatattgtctataaacaaaactaaag
Dana   725  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dvir   889  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dmoj   735  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dgri   835  gtttaaaattatatatgttgcgccgcacagctgaagattttccttagagtgaagacaagc
```

```
Dmel   1089 ~aac~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~a~~~~~~~~~~~~aacagtc~~~aa
Dsim   1098 ~aac~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~a~~~~~~~~~~~~aaccgtc~~~aa
Dyak   1084 ~taacctttatactaaaaagatgaaacattggaaattatttaaaatgtgaa~gttttttaa
Dere   939  gtaaccgctatactaaaatgatgaaacattcgaaatttt~~~~~~~~gaa~~gttttttaa
Dpse   1008 ttttactgatggtgtcttaaacacggaactgttattgagaaagtcttggaccatattctg
Dper   1066 ttttactgatgctgtcttaaacacggaactgttattgagaaagtcttggaccatattctg
Dana   725  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dvir   889  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dmoj   735  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dgri   895  tcatgtgaaatgaaatggcttagtttgggcttagatattcttgttattttgtgaactgtg

Dmel   1103 tatttctttag~aggattagttaaatcatcacatttcagata~catacatttatttattt
Dsim   1112 tatttctttat~aagatta~~~~~~~~~~~~~~~~~~agata~catacatttatttattt
Dyak   1143 tattt~tttatcaa~a~c~~~~~~~~~a~~~~~~~~~ttct~~~at~~atttttttata~
Dere   989  tattt~ttca~~acgaac~~~~~~~~~~~~~~~~~~~attcagcat~~acttctttataa
Dpse   1068 aagaagcgaagccctctaaccacaaacttgtttccaattggcgtgaattttcttc~~~~~
Dper   1126 aagaagcgaaaccctctaaccacaaacttgtttccaattggcgtgaattttcttccaaaa
Dana   725  ~~~~~~~~~~~~~~t~t~~~~~c~a~~tttatt~~~att~~~~tg~att~~~ttc~~~~~
Dvir   889  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dmoj   735  ~~~~~~~~~~ccgaaaactcaacattattcgagaatatttcaaatattgtttcaaaacaa
Dgri   955  atttgtgtcttgcagttcaataa~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

Dmel   1160 ~ttgtatatatgtatat~atat~~~t~~tttg~~~gtcagctcatttaattta~~~~~ca
Dsim   1151 ~atg~~~~~~~~tacat~atat~~~a~~tttg~~~gtcagctcatttaattta~~~~~ca
Dyak   1175 ~agaaaaa~~t~taaatcat~~~~cacatt~~~~~~ccagctcatttatttaattttca
Dere   1025 gat~tagt~~~~tgaatcacattccagctttgtatgt~~~ctcgtttaatttaattttta
Dpse   1122 ~~~~~~~~~~~~~~~~~~atcatttcaatgatgaatcacttaagataatggattcaaaag
Dper   1186 tgaataaagtatgaacatagcatttcaatgatgaatcacataagataatggattcaaaag
Dana   746  ~~~~~~~~~~~~~~~~~~attaatttaat~~taaat~ac~~~agataatgaaat~aagag
Dvir   889  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dmoj   786  ttttaaggctgtataagaaataatccattaacgttcatctaaagatagtttttaaaaatt
Dgri   977  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

Dmel   1206 gttctttcaccgtttggtatc~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dsim   1189 gttctttcaccttttggtatc~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dyak   1222 gttcatacaccttttggtatc~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dere   1077 gttctttcgccttttgggttc~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dpse   1165 aactcctaaaaat~a~tata~~t~atagga~cattactatagccctacctcctggcttcg
Dper   1246 aactcctaaaaat~a~tata~~t~atagga~cattactatagccctacctcctggcttcg
Dana   781  ~atttctaaaaatgactttaaattatagcctcatt~gtata~~~~t~~~t~ct~~~tt~g
Dvir   889  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dmoj   846  cgttaataaataatgttaatattcattaaaaaataattctaaaaaaataataattcttt
Dgri   977  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

Dmel   1226 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dsim   1209 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dyak   1242 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dere   1097 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dpse   1219 t~~ttctgtgccataaaaagagacactt~t~aacttcttgccaactcatt~tt~tttt~a
Dper   1300 c~~ttctgtgccataaaaaaagacactt~t~aacttcttgccaactcatt~tt~tttt~a
Dana   828  taatt~t~t~~~at~~gca~agtc~ctcatcaaagtcatgacagctcagtcttattttga
Dvir   889  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dmoj   906  cagcggaattcgagaatagttaagacatactaagtagatcaagaatacttaaataaaact
Dgri   977  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
```

```
Dmel     1226 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ctgtttgtctgcaaactgttttaccg~~~t
Dsim     1209 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ctgtttgtctgtaaactgttttaccg~~~t
Dyak     1242 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ctgtttgactgttaactgttctactg~~~t
Dere     1097 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ctctttgtctgtaaactgttttaaaaaaat
Dpse     1272 agttccttcaacttgt~~tttgtt~ttt~tctgtttatctgttaagcattttaccg~~~t
Dper     1353 agttccttcaacttgt~~tttgtt~ttt~tctgtttgtctgttaagcattttaccg~~~t
Dana     879  agttctttcacctt~taatttgttgtttgtct~tttgtctgtaaaccgctttaccg~~~t
Dvir     889  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dmoj     966  tcctagttcttattagacattcaaataaattgtttctgcctagttgaacaatcgttccag
Dgri     977  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

                                                                        --
Dmel     1253 ~~tgcattcttgcaaatataaaaaa~ct~~t~~~~~~~~~~tttcgacttg~~~~aagtg
Dsim     1236 ~~tgcattcttgcaaatataaaaaact~~t~~~~~~~~~~tttcgactcg~~~~gagtg
Dyak     1270 aatacactcttgc~aatat~~~a~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dere     1127 ~~ttcagt~tt~~~aa~at~~cact~ctgat~~~~~~~~~~tttg~~t~~~~~~~a~tg
Dpse     1324 ~~tatatccgtgcaaataatatacaaagaaaccacacaacctttcgacttgg~~~aagtg
Dper     1405 ~~tatatccgtgcaaataatatacaaagaaacagcacaacctttcgacttgg~~~aagtg
Dana     933  ~~tatatccgcgcaaata~~~tacaaa~~~~~~ac~ca~~~~~t~tttgtcttaaaaatg
Dvir     889  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dmoj     1026 actggttcaaattcaaattcagttcaagtgcagctgccttacgttgaatgtgccttgagc
Dgri     977  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

               AFB
              -----
Dmel     1295 aatcactc~ggatttcggtaaattat~~~~~~~tttctacccaactttaccgtaaccaaa
Dsim     1279 aatcactctgattttggtgaattcat~~~~~~~tttctacccaactttgacttaatcaat
Dyak     1288 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~gat
Dere     1157 ~~~~~~~~~~~~~~~~~~taaattttgc~cgt~tttctgctctacccg~acctaaataat
Dpse     1380 ca~gcaa~aatgatatgtgactcaccaaaacattttttgggtac~~~~~~~~~~~~~~aa
Dper     1461 ca~gcaa~aatgatatgtgactcaccaaaacattttttgggtac~~~~~~~~~~~~~~aa
Dana     976  ca~acaa~~~~~agctccg~~tc~~~~acacagtttttgg~~~~~~~~~~~~~~~~~~aa
Dvir     889  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dmoj     1086 tttatctgccaattctactcagtgctcgcagcacacctcaacat~~~~~~~~~~~~~~cc
Dgri     977  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~accttaaaga~~~~~~~~~~~~~~gt

Dmel     1347 caaaaggcgttcagttgcagaaaacttgactattttttcaataatgatgctcaagattga
Dsim     1332 taaaagtcattcaattgcagaaaacttgac~~tgttttcaataatgatgtataagagtga
Dyak     1292 t~~aa~~~~~~~~~~~~~a~~~~~~tt~~~~~~~~~~~~~~caggta~~~~~~~~~~~~~
Dere     1197 taaaaggcattcaattgcagacaacttaaa~~tgttttcaataatgatgtagaaggttga
Dpse     1424 tacaagaga~gtttttc~~~~~~~~~~~~~~~~~~~gttttcagccaattaaaaggcatt
Dper     1505 tacaagaga~gtttttc~~agttt~acaaatttttcgttttcagccaattaaaaggcatt
Dana     1006 tataattggcgttttggcctgtttgac~ga~~~~~~~~~ctaaaccaattaaatggcatt
Dvir     889  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~T
Dmoj     1132 actcaagcgctgtcaattcaaaattcttaagcttgatttttttctttttgtatattcatt
Dgri     990  tccttaaag~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

Dmel     1407 atcgtggtattttgatt~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dsim     1390 atc~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dyak     1303 ~~~~~catattttgatt~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dere     1255 atcaaggtattttgatt~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dpse     1464 aaaatgcaagaatta~~~~~~~~~~caatg~~ctattgagggattcatgtggca~ct~cg
Dper     1561 aaaatgcaagaatta~~~~~~~~~~caatg~~ctattgagggattcatgtggca~ct~cg
Dana     1056 aaattgcaaaaaatatgattgtctacattgggct~ttga~~aattcaaaaagcatctgca
Dvir     891  TCGCCAACTCATTGAGTCAAGTTCCAGTCACTTATTATCTGCCATCATAATACTGTTATG
Dmoj     1192 ttgccaactcattgagtcaagttccagtctcgtattatttaccaacgtattaccgttatg
Dgri     998  ~~~~~~~ctcattaagtcaagttggcttgccatc~~~~~~~~~~~~~~~~~~~gttata
```

```
Dmel    1423 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~gtgggaa~~~~~~
Dsim    1392 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~gtgggaa~~~~~~
Dyak    1315 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~atgtgga~~~~~~
Dere    1271 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~atgtgaa~~~~~~
Dpse    1510 gat~~gtt~~g~tgcatctataaactgatttgaagtgattttaacaagtgtgaa~~~~~~
Dper    1607 gat~~gtt~~g~tgcatctataaactgattcgaagtgattttaacaagtgtgaa~~~~~~
Dana    1113 gatcagtcaagatgaat~t~taatttaaatcgaa~tcattttgatta~tatgaa~~~~~~
Dvir    951  CAGATACACATAAATAAATTCAAGACCTCAACCAACT~~~~~~~~~~~~~~~~~~~~~~~
Dmoj    1252 cacatgc~~~~~~ataaattcaagaccccagccaacaacgt~~~~~~~~~~~~~gtgagt
Dgri    1033 cacattcacataaataaatacaagacctccatttgagt~~~~~~~~~~~~~~~~ctgagt

Dmel    1430 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dsim    1399 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dyak    1322 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dere    1278 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dpse    1558 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dper    1655 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dana    1162 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dvir    987  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dmoj    1293 cacgaagattttgtacga~~~~~~~~~~~~ttcataaactttgctcataactttgacagc
Dgri    1077 cacaagttttcggcggttctctttgcattcttcatata~~~tgctcatttccttacaccc

Dmel    1430 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dsim    1399 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dyak    1322 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dere    1278 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dpse    1558 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dper    1655 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dana    1162 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dvir    987  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dmoj    1341 attttagtgcaat~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dgri    1134 ttgcaatgggtattataatttcctggtgaggtttgttatttattgaaggcaacgtttctg

Dmel    1430 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dsim    1399 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dyak    1322 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dere    1278 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dpse    1558 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dper    1655 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dana    1162 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dvir    987  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dmoj    1353 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dgri    1194 ggaacatcttttgttactttgttacttaacactcaatgctcataaaatttggagacgtac

Dmel    1430 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dsim    1399 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dyak    1322 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dere    1278 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dpse    1558 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dper    1655 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dana    1162 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dvir    987  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dmoj    1353 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dgri    1254 aaaatcactcagactactcaatgatatagctttcataagaacataagtagaaagtcagat
```

```
Dmel   1430 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dsim   1399 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dyak   1322 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dere   1278 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dpse   1558 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dper   1655 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dana   1162 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dvir   987  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dmoj   1353 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dgri   1314 tttggcagattgcctctactctgcaaaggtttttaatcttcgggatgccaaagtgttgtt

Dmel   1430 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dsim   1399 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dyak   1322 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dere   1278 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dpse   1558 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dper   1655 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dana   1162 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dvir   987  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dmoj   1353 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dgri   1374 atttttgataattagtatgtctcaattaaactgacggaaatcaataggaattcttggtt

Dmel   1430 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dsim   1399 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dyak   1322 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dere   1278 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dpse   1558 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dper   1655 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dana   1162 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dvir   987  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dmoj   1353 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dgri   1434 aacatcaatgattctacgaaacttcgatacatatgtatatatgcatatatacattgttca

Dmel   1430 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dsim   1399 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dyak   1322 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dere   1278 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dpse   1558 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dper   1655 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dana   1162 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dvir   987  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dmoj   1354 aatctgaaacatcattaaggattttcggctctatacaatagagttatattaaggtatatt
Dgri   1494 ta~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

                              AFB
                            -------
Dmel   1430 ~~~~~~~~~~~~~~atgagtcacaaaagagctt~aatatttggattttacgtaatgagtg
Dsim   1399 ~~~~~~~~~~~~~~atgagtcacagaagagctt~aatatttggattttacgtaatgagtg
Dyak   1322 ~~~~~~~~~~~~~~atgagtcacagtagtgctt~aatatttggattttacgtaatgactg
Dere   1278 ~~~~~~~~~~~~~~atgagtcacagtagaggtt~aatatttggattttacgtaacgagtg
Dpse   1558 ~~~~~~~~~~~~~~atgagtcacagtagagctt~aatgtttggttttttacgttatgcgta
Dper   1655 ~~~~~~~~~~~~~~atgagtcacagtagagctt~aatgtttggttttttacgttatgcgta
Dana   1162 ~~~~~~~~~~~~~~gtgagtcactg~aaaactttaatgtttggtttttacg~ta~aagta
Dvir   987  ~~~~~~~~~~~~~~CTGACTCACAAAATTTGAACGGCTCGTACGCCTTCCACTGAATTCC
Dmoj   1414 tagtctgttgcgctatgagtcacaagagtttgcacaaaact~~~~~~~~~~~~~~~~~~~
Dgri   1495 ~~~~~~~~~~~~~~atgagtcacgaa~gtttgctcataac~~~~~~~~~~~~~~~~~~~~
```

```
Dmel   1476 agttgaa~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dsim   1445 agttgaa~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dyak   1368 atttgaa~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dere   1324 agataaa~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dpse   1604 agtcg~a~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dper   1701 agtcg~a~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dana   1206 agccgga~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dvir   1034 ATCCGTACTGAATGCCTTTTTTGCCCATAACTAAGGACTCTCCATACAATATGCATATGT
Dmoj   1454 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dgri   1520 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

Dmel   1482 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~tcggaactggc
Dsim   1451 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~tcggaactggc
Dyak   1374 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~tcggactggcc
Dere   1330 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~gggactgac
Dpse   1609 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~atatcatcttt
Dper   1706 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~atatcatcttt
Dana   1212 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~at~t~gtcttt
Dvir   1094 ACTTTAATTATGCACCATGATTCACAAAGTTTACTCATAACCAAAAAAG~~~~~~~~~~~
Dmoj   1454 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dgri   1520 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ccagatag~~~~~~~~~~~~~~~~~

Dmel   1494 acttcatttttgggtttttctttttttttttttggtttcccttcagacctgcgaatgatt
Dsim   1463 acttcatttttgggtttttc~~~~ttttctttcggtttctctgcagacctgcgaatgatt
Dyak   1386 ctttaatttttgggtt~~~~~~~~~~~~~ttttggTTTCTTTTCAGACCGGTGAATGATT
Dere   1340 acttcatttttggg~~~~~~~~~~~ttt~ttttggtttcttttcagacctgagaatgatt
Dpse   1620 ~c~gtgctcttgtgcactcgtgg~t~~~~~~~~~~~~~~~~~~~~~~~ctgagaatgatt
Dper   1717 ~c~gtgctcttgtgcattcgtgg~t~~~~~~~~~~~~~~~~~~~~~~~ctgagaatgatt
Dana   1222 tcagt~ttctt~t~ttttcggggat~~~~~~~~~~~~~~~~~~~~~~~ctgagaatgatt
Dvir   1142 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dmoj   1454 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dgri   1528 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

Dmel   1554 tgtct~tt~tgttc~~~~~~~~~~~~~~~~gggctttaatttgagcg~accattgat~~~
Dsim   1519 tgtct~tt~tgttc~~~~~~~~~~~~~~~~gggctttaatttgagcg~accattgat~~~
Dyak   1434 TGTCT~TT~TGTTT~~~~~~~~~~~~~~~~GGGCTTTAATTTGAGCG~CCCATTGAT~~~
Dere   1388 tgtct~tt~tgttc~~~~~~~~~~~~~~~~gggctttaatttgagc~accaattgct~~~
Dpse   1655 tgtct~ct~tgtttggattgggggg~actgggggctttcatt~~~~~~~~~~~~~~~~~~~
Dper   1752 tgtct~ct~tgtttggattgggggggactgggggctttcatt~~~~~~~~~~~~~~~~~~~
Dana   1256 ggtcttctgtgagtagatt~~cgg~~~ttggggttttcgtt~~~~~~~~~~~~~~~~~~~
Dvir   1142 ~~~~~~~~~~~~~~~~~~~~~~~~TGGTTTTTACGCTATGCGTAACTTAATTTTTATCTG
Dmoj   1454 ~~~~~~~~~~~~~~~~~~~~~~~~tggtttttacgcagtgcgtaacttaattcgaaactg
Dgri   1528 ~~~~~~~~~~~~~~~~~~~~~~~~tgatttttacacaatgcgtaactgaattcaaagctg

Dmel   1591 ~~~~gggtttattttggg~ttttggaagacttattcatt~~~~catc~~~~~~ga~~tga
Dsim   1556 ~~~~gggtttattttggggttttggaagacttatccatt~~~~cacc~~~~~~ga~~tga
Dyak   1471 ~~~~GGGTTTATTTTT~GGTTTTGGAAGAATTATCCATT~~~~GACC~~~~~~GA~~TGA
Dere   1425 ~~~~ggtttattttt~gggttttggaagacttagccatt~~~~cacc~~~~~~ga~~tga
Dpse   1692 ~~~~gatgactaaa~~~~~~~ttgaaggacttattcatg~~~~aatccc~~ctga~~tga
Dper   1790 ~~~~gatgactaaa~~~~~~~ttgaaggacttattcatg~~~~aatccc~~ctga~~tga
Dana   1291 ~~~~gatgactgta~~~~~ttttgga~~~~~tattgaaggaccaatccaagctcagctga
Dvir   1179 GTGTTAAGCTTTGGGCTTCACCCATTACCAATTAACATTAAAGAGATAAATTATATTTTG
Dmoj   1491 gtattaaattatgggatttttttttatttttttttactcattgaaatgaggtgaatcggc
Dgri   1565 tcatttaatttgggttttttttgtatacaatattattcatctctgttttctgcatgtttc
```

```
Dmel   1635 gtt~gtcatgttt~~aatgacacac~ttat~ttt~~~ggtcataa~g~~~~~~~~~~~~~
Dsim   1601 gtt~gtcatgttt~~aatgacacac~tttt~ttt~~~ggtctt~agg~~~~~~~~~~~~~
Dyak   1515 GTT~gtcatgttc~~aatgacacac~tttt~ttt~~~ggg~~~c~~~~~~~~~~~~~~~~
Dere   1469 gtt~gtcatgttt~~aatgacacacattttcttttttgg~cttca~g~~~~~~~~~~~~~
Dpse   1734 gtt~gtcaaaatg~~aattatttaacatagtta~~~~~~tcatagtg~~~~~~~~~~~~~
Dper   1832 gtt~gtcaaaatt~~aattatttaacatagtta~~~~~~tcatagtg~~~~~~~~~~~~~
Dana   1338 gtt~gaaattttt~~aatgatataatatagttaatgaatcagatagc~~~~~~~~~~~~~
Dvir   1239 ATTATTGGATAGCTCCTTTGCATATGAACGGTTTTCGCCTTTTCAACTTTACGTAAAAAA
Dmoj   1551 tcataatgaagctgaacttaagttaaagtggaagaaggtttgctcaggctgtgggaacca
Dgri   1625 tcgcttaaataagcttgtctttaaattgagggcttggtttttttaaagttattgaaggac

Dmel   1672 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~cactttgaaactt~~~~~gga
Dsim   1638 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~cttggagaacttt~~~~~~~~
Dyak   1547 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~acattttatgcatt
Dere   1510 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~cttgg~gacattgtgtggttt
Dpse   1771 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~a~~~~~~aat~atgatttagc
Dper   1869 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~a~~~~~~aat~atgatttagc
Dana   1381 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~agggagaaatcaagttttaaa
Dvir   1299 AATCAAAAACAGTTATCAGTCCAAATCTGTCCGATTTGAGC~~~~~~~~~~~~~~~~~~~
Dmoj   1611 taaaagagcctaaatgaaatgctcagtataaagtctaacat~~~~~~~~~~~~~~~~~~~
Dgri   1685 aagctaaataaaaatagacttatgaagttgaataataaata~~~~~~~~~~~~~~~~~~~

Dmel   1689 atttaaaggcatttaagg~~~~~catgttaaagaaaa~~~~gttatagccc~~gctcaat
Dsim   1652 gtgacttcgaatttaaagtaagataagttcaagaaaa~~~~gttatagccc~~gctcaat
Dyak   1562 g~gattttaaatct~~~~g~~~~~~~~~~~~~~~~aagttggatat~gtacaag~t~aat
Dere   1531 g~gattttgaatctgaagtaagatatgtaaaacaaaa~~~~gttatatcct~~gctcaat
Dpse   1786 cttgaa~attggtaaaaggtgttcccccacca~attt~ca~~aca~t~~caa~~~~~tta
Dper   1884 cttgaa~attggtaaaaggtgttcccccacca~attt~ca~~aca~t~~caa~~~~~tta
Dana   1403 aatgcatatt~ttaaaacctctttctata~gagattttcagga~agtttcatcgatttta
Dvir   1339 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dmoj   1651 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dgri   1725 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

                                           GRH                       GRH
                                         -------                    ----
Dmel   1738 gc~aaaaacgttctttc~aaagatttaagactagtttttaaaaacagttcgatcaaacc
Dsim   1706 gc~aaaaacgttcttttt~aaagatttacgacaag~~tttaaaaaaaaattccaataaacc
Dyak   1598 gccaaaaaagat~tttt~aaatatttagacttg~~tttta~~~~~aatttgaataaacc
Dere   1584 gc~aaaa~agtactttttaaatatttagactag~~ttttataaacaatttgaatctacc
Dpse   1833 t~gt~~~ttcgtttttaacgatgt~ttga~g~~~~~t~g~~tattt~~tt~gaa~~gc~~t
Dper   1931 t~gt~~~ttcgtttttaacgatgt~ttga~g~~~~~t~g~~tattt~~tt~gaa~~gc~~t
Dana   1460 ttgtatttttttttttaagtatttattaatgataaatagcttgtttccttagcattgactt
Dvir   1339 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dmoj   1651 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dgri   1725 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

            ----
Dmel   1796 gctttggcattgactttgaagttg~ttg~~~gtctggc~~gtac~ctctgcaattgaat~
Dsim   1762 gctttagcattgactttgaagttg~ttg~~~agctggc~~ttac~ctcttcaattgaat~
Dyak   1649 acgttagcattgaatttgaaatttattcttaaactggcgatttaaactcttcgatt~aata
Dere   1640 cctttagcattgactttgaca~~~~~~~~~~~~~~~~~~~tt~~~ctcttc~gtttacta
Dpse   1872 taaagtg~~a~~gaga~~~~~~~~~~aaat~~~~aaa~cgtttg~~ctggtgtt~~ttg~
Dper   1970 taaagtg~~a~~gaga~~~~~~~~~~aaat~~~~aaa~cgtttg~~ctgatgct~~ttt~
Dana   1520 taaatttcaactga~atgtttactttaatttttaatgc~ttgaaacaacagttacttgc
Dvir   1339 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dmoj   1651 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dgri   1725 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
```

```
Dmel    1848 actccaaacaacagttacctgctacat~~~~~~~~~~caaatcaacattgactatttgct
Dsim    1814 actccaaacaacagttacctgctacattgagagaaaacaaataaacattgactatttgca
Dyak    1708 actccaaacaacagttacctgctata~tg~gag~aaacaaattaacattgactatttgca
Dere    1677 actccaaacaacagttacctgctatatggatagaaaacaaataaacattgacaatttgca
Dpse    1907 ~~~gttgg~~gc~~~~~~~~~~~~~~~~~~~~~~~~~~~~~tg~~~~tgggg~~~~aaga
Dper    2005 ~~~gttgg~~gc~~~~~~~~~~~~~~~~~~~~~~~~~~~~~tg~~~~tgggg~~~~aaga
Dana    1578 catgttgacagaaacct~~~~~~~~~~~~~~~~~~~~~~~ataaacattgagccacaa~a
Dvir    1339 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~AATAAACATTGAATATTTGCG
Dmoj    1651 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dgri    1725 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~aaagcttaataattttgtctt

Dmel    1898 aagccttttgtttggcaacgcgccagtggaagaatgtgagcgaaattcaaatgaccaagc
Dsim    1874 gaggcttttgtttggcagcgcgccactggaagaatgtgagcgaaatttaaatgagcaagc
Dyak    1765 gaggcctt~~att~~cggcgcgccaat~caaaaatgtgagcgaaatttaaatgagcaaga
Dere    1737 gaggcttttatttggcagcgctccactcgaaaaatgtgagcgaaatttaaatgagcaagt
Dpse    1926 a~~~~~~~~~~~~~ga~~~~~g~~taaacagc~agatgaagaagcg~agagatac~a~ac
Dper    2024 a~~~~~~~~~~~~~ga~~~~~g~~taaacagc~agatgaagaagcg~agagatac~a~ac
Dana    1615 a~~~~~~~~~~~~~gatttgtgagtaaa~atttaaatg~agaggcgca~aaatacgagaa
Dvir    1361 GAGACATTTTAGCTGAGCGCAAGAGAGACG~~~~~TTGTCTAAAATT~~~~~~~~TAAAT
Dmoj    1651 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dgri    1747 gattgtaagtagtgttaatatgtctataaattatgggatgctatttttttatttactttt

Dmel    1958 tg~cgggagaaaaatcga~gagaagagacc~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dsim    1934 ta~cgggagaagaatcga~gagaagagacc~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dyak    1820 ag~tgggagaagaatcga~gagaagagacc~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dere    1797 ag~cgggagaagaatcat~gagaagagacc~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dpse    1961 ~aaagag~~~~aga~~cc~~gagc~agact~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dper    2059 ~aaagag~~~~aga~~cc~~gagc~agacc~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dana    1659 gaaacagctagaaaggcgaagagccagacc~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dvir    1408 AAGAAAGACAAACAGT~A~GAGAAGAGAGC~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dmoj    1651 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dgri    1807 aaataccaagttttatgctttcgaactgtggttacttgaaaaaacagttacttgtgctat

Dmel    1985 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dsim    1961 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dyak    1847 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dere    1824 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dpse    1981 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dper    2079 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dana    1688 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dvir    1435 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dmoj    1651 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ataagatataaaa
Dgri    1867 attgacaaaaaaaaaatcgtcataaaaaggcgtacgaataaggcaaaaataaacattaaa

Dmel    1985 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dsim    1961 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dyak    1847 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dere    1824 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dpse    1981 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dper    2079 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dana    1688 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dvir    1435 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dmoj    1665 ccattttatatatgtaacaatttagtttaattcagaatgttcacatacaaaatctttcta
Dgri    1927 tatttgccaagacactttggcagctgctgagcgagacattgcccaaaatataaatgagtt
```

```
Dmel    1985 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~tcgagactggagccttgaagac
Dsim    1961 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~tcgagactggagccttgaagac
Dyak    1847 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~tc~~~~~~~gacttttgaggac
Dere    1824 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~tc~~~~~~~gagccttgaagac
Dpse    1981 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~tcgagactgtgagattccgg~t
Dper    2079 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~tcgagactgtgagattgcgg~t
Dana    1688 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~tcgagacagagaggtt~tggct
Dvir    1435 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dmoj    1725 tgtttttctatacatttcttaagcttccataaaagattttagtaaacaaactttgtctat
Dgri    1987 aaac~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

Dmel    2008 agaatctcaaagcctgagtgc~~atctcgc~~~~~~~~~~~~~~~~~~~~tgg~~~~~~~
Dsim    1984 agaatctcgaagactgagcgc~~atctcgc~~~~~~~~~~~~~~~~~~~~tgg~~~~~~~
Dyak    1863 agaatcttgcagactgagcgcgcatctcgc~~~~~~~~~~~~~~~~~~~~tgg~~~~~~~
Dere    1840 agaatcttaaagacggagcgc~~atctcgc~~~~~~~~~~~~~~~~~~~~tgg~~~~~~~
Dpse    2002 ~~~~~~~~~ggaggtggaggt~~~~~ggaggcgtgtcaacaaa~tgtgattgg~~~~~~~
Dper    2100 ~~~~~~~~~ggaggtggaggc~~~~~ggaggcgtgtcaacaaa~tgtgattgg~~~~~~~
Dana    1710 aaacaactcgga~ct~gaggctaagtggaggcgtttgagcaaaatccgattgg~~~~~~~
Dvir    1435 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~CAAAAAAAAATATATG~GAGAGC
Dmoj    1785 ttatgcaaagattttgacagtcgctgcgagaagagtaaagtaaataatagaaaagagagc
Dgri    1990 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~aaacaagagagaagagatt

Dmel    2038 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dsim    2014 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dyak    1895 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dere    1870 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dpse    2040 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dper    2138 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dana    1760 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dvir    1457 GAGAGAGAGAGAGAGAGAGGCCGACTCACTGCTTGTATATCTCCACAAAGAAAAAAACAT
Dmoj    1845 gaaaaagagagagagaaagagagcaatcgtatttacccagcaacaagctactcagacaac
Dgri    2010 gagaaccatagacacagagagtcggagaggcttgtaaatcaccatataaataagcattaa

Dmel    2038 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dsim    2014 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dyak    1895 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dere    1870 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dpse    2040 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dper    2138 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dana    1760 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dvir    1517 CGTCTTGGTTT~~~~~~~~~~~~~~~~~~~~~CTTGGCTTGCATGTAGGCGGTTCGGCCT
Dmoj    1905 tctgtatctctatgg~~~~~~~~~~~~~~~~~cttggcttgcgtgtgggcgttgaggcct
Dgri    2070 agccaactcttcgacagtgtctcttgaatttgttgtatt~~~~~~~~ggcgttgaggcat

Dmel    2038 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~aaa~tgccaagcggatgagtagacg
Dsim    2014 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~aaa~tgccaagcggatgagtggacg
Dyak    1895 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~aaa~tgccaagcggatgagcggacg
Dere    1870 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~aaa~tgccaagcggatgagtggatg
Dpse    2040 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~aaa~tgccaaacagatgatttggca
Dper    2138 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~aaa~tgccaaacagatgatttggca
Dana    1760 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~aaa~tgccaagcggatga~gtg~~~
Dvir    1556 TTAAAATTATGCGAAAATTGACTGGCACAGATTAAAAA~TGCCAATTAGATGAACGCCCT
Dmoj    1948 ttcgaattgtgcaaatattttgagtggcaaacatgaaa~gagtcaattggaaaaagttgt
Dgri    2122 ttaagattgtgtaaaactcaagtggcatggattagaaaatgccaattagatgaaactttt
```

```
Dmel   2063 tactcg~~~~~~~~~~~~~gctgagccaggtgaaaaccctcgctaggtgacctttcccgc
Dsim   2039 tactcg~~~~~~~~~~~~~gtggagccaggtgaaaaccttcgccaggtgacctttctcgc
Dyak   1920 tactcgt~~~aggtaatgagctgagccaggtgaaaaccgtcgtcaggtgaccttttccgc
Dere   1895 tacttgg~~~cagtaaagagcttagccaggtgaaaaccgtcgccaggtgacctttcccgc
Dpse   2065 gctat~atacgagta~~catatctacatatgtatgatctaga~gaagatcttggg~a~~t
Dper   2163 gctat~atacgagta~~catatctacatatgtatgatctaga~gaagatcttggg~a~~t
Dana   1781 g~~atgat~~gagccgcca~at~tgga~a~g~atgatccaggtgaaaacccacggcaggt
Dvir   1615 GGATACTTTTAGTAAG~~~~~~~GATTTGATTTAGTGAATTTTAACCGTTGATGCCAGCA
Dmoj   2007 attaaatgtcatttattccaaatgatttcatttggctcatttcacttaataacaggctta
Dgri   2182 ctctactgacattttgggcgaatttcacccatcgatattaggtgacggcatagttaacag

Dmel   2110 aggtaattc~~~~~~~~~~~~cc~tgtgaaagcctc~~~~caaaaacc~attgcgcagcc
Dsim   2086 gggtaattc~~~~~~~~~~~~cc~tgtgaaagcctc~~~~caaaaaaccactgcgcagcc
Dyak   1977 tggtaattc~~~~~~~~~~~~ca~tgtgaaagcctt~~~~cgaaaaca~actgagctgca
Dere   1952 tcgtaattc~~~~~~~~~~~~cc~tgtgaaagcctc~~~~caaacacc~actgtgttgca
Dpse   2118 c~~~~~~~~~~~~~~~~~tt~cc~tgagaaagaagcccagcaaaaaccccacaaagtgtt
Dper   2216 gaaaataagcgt~~~~~atttcc~tgagaaagaagcttagcaaaaaccccacaaagtgtt
Dana   1832 gacatttcgctttggcgatt~caatgtgaaagtatcc~~tcaaaaaa~~~~~~~~~tg~t
Dvir   1668 ACAGCTACTCCATCTATATAGTTAAATGTTTT~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dmoj   2067 gccatagcatcga~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dgri   2242 ctctctgttaaat~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

Dmel   2152 cagaaatcttggcactgactctcctctgctga~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dsim   2129 cagaaatcttggcactgactctcctctgctga~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dyak   2019 aagaaatcttaacgctgattctcctctgctga~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dere   1994 aagaaatatgggcgctggttctcttttgctga~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dpse   2159 agc~cag~~t~~~~~~~~~~~~~ctctgctaaatgccaagaaaatggtttgacttgagat
Dper   2270 agc~cag~~t~~~~~~~~~~~~~ctctgctaaatgccaagaaaatggtttgacttgagat
Dana   1879 acctcatgat~~~~~~~~~~~~~ctccgttaaatgccaagagaatgctgccacttggctc
Dvir   1699 ~~~~~~~~~~~~~~TTAGTAATGCTCTGCTAATTAAGG~~~~~~~~~~~~~~~~~~~~~~
Dmoj   2079 ~~accatag~~~~~ttaacaagtctttgttaattaaga~~~~~~~~~~~~~~~~~~~~~~
Dgri   2254 ~~gccaagaa~~~~~~~~~aatggcttgctaattaagg~~~~~~~~~~~~~~~~~~~~~~

                                                 GRH
                                               --------
Dmel   2183 ~~~~~~~~~~~ctaatttaatc~~ataatttatagaacaagttgctgaatttt~gggtgt
Dsim   2160 ~~~~~~~~~~~ctaatttaatc~~ataatttatagaacaagttgctgagattt~gggtgt
Dyak   2050 ~~~~~~~~~~~ctaatttaatc~~ataatttattgaacaagttgctg~gattttgggtgt
Dere   2025 ~~~~~~~~~~~ctaatttaatc~~~taagttatagaacaagttgc~gagattttgggtgt
Dpse   2203 gccttg~~~~~ctaatttaatttttataatttatagaacaagttgctcggttttt~gg~gt
Dper   2314 gccttg~~~~~ctaatttaatttttataatttatagaacaagttgctcggttttt~gg~gt
Dana   1926 gccttggttagctaatttaatttttataatttatagaacaagttgcttggttttttggtgt
Dvir   1723 ~~~~~~~~~~~~~~~~~~~~~TTTATAATTTATAGAACAACTTGTTTTAAAGGTGTGTTT
Dmoj   2110 ~~~~~~~~~~~~~~~~~~~~~ctttcaatttatagaacaacttgttttaaaggcgtctgt
Dgri   2281 ~~~~~~~~~~~~~~~~~~~~~tttatgatttatagaacaacttgtttttaaggtgtgctt

Dmel   2230 gtt
Dsim   2207 gtt
Dyak   2097 gtg
Dere   2071 gtt
Dpse   2256 gtg
Dper   2367 gtg
Dana   1986 gtg
Dvir   1763 GTG
Dmoj   2150 gtg
Dgri   2321 gtg
```

msn1.2SubB

```
Dmel   1     CCACTGC~~~~CAACAGCAAGAA~~~~~~~~~~~~ATCAGTGCACTTGCCATAAGACCGT
Dsim   1     CAACTGC~~~~CAACAGCAAGAA~~~~~~~~~~~~ATCAGTGCACTTGCCATAAGACCGT
Dsec   1     caactgc~~~~caacagcaagaa~~~~~~~~~~~~atcagtgcacttgccataagaccgt
Dere   1     ccacagcca~~~cacagctaaaa~~~~~~~~~~~~atcagtgcacttggcataagtccgt
Dyak   1     ccactgcacccccacagcaaaaaaaaaaaaacaaaatcagtgcacttggcataagtccgt
Dana   1     TTCTGGTTCTGGTGCCCTGTCTGCCCTCTTGACAGTTTCCA~GTTGACCACGTTT~CGAG
Dsec   1     CACTCTGTCTCGCTATCTCTCTCTCTTTCTCT~~~~~~~~ACCCGTAGAATCAAAGCCGT
Dpse   1     cactctctctctctctctctctctctctctctttctctacccgtagaatcaaagccgt
Dvir   1     ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dmoj   1     atccaatcagagcatttgcccacaccccgctgtttgcttttaccgttaacagtaaa~~~~
Dgri   1     acaaccacaatacaatacatacaaagctctcccacgggggactccacctccggagactac

Dmel   45    TAACACGTTGCACTTGTACGTGTTCCCCGGAACCG~~CATGTGGCTAACGATCCGATCCT
Dsim   45    TAACACGTTGCACTTGTACGTGTTCCCCGGAACCG~~CATGTGGCTAACGATCCGATCCT
Dsec   45    taacacgttgcacttgtacgtgttccccggaaccg~~cgtgtggctaacgatccgatcct
Dere   46    taacacgttgcacttgtacgtgttccccggaaccg~~catgtggctaacgatcc~~tcc~
Dyak   61    taacacgttgcacttgtacgtgttccccggaaccg~~catgtggctaacgatcc~~tcct
Dana   59    TCAGACTgtgcacttgtacgtgttctgtggaagca~~catgtgcccaacgcccatctgaa
Dsec   53    TAAAACGTTGCACTTGTACGTGTTCTACAGGGAGCCACATGTGGTGTTCCTCCCTCATAA
Dpse   61    taaaacgttgcacttgtacgtgttctacagggagccacatgtggtgttcctccctcataa
Dvir   1     ~~~~~~~~~~~~~~catacacacactcacacacacacacacctacacacacatgcatgca
Dmoj   56    ~tgttccggattaacatacacatctacatacgacaacacacatacacacacacaccttcg
Dgri   61    gtgttccagattcacatgtacaaccaaacctatatttttacatgtgttcatcaagctctc

Dmel   103   ACC~CCCACCCACG~ACTCCTCCGCCACGATTCCATTCCAATTCAGCTCCT~~~~~~~~C
Dsim   103   CCC~CTCGCCCACG~ACTCCTCCGCCACGATTCCATTCCAATTCAGCTCCT~~~~~~~~C
Dsec   103   ccc~ctcacccacg~actcctccgccacgattccattccaattcagctcctc~tgtcctc
Dere   101   cccacttcgcc~~~~actcctccgcagcgattcccttccatttcagctcct~~~~~~~~c
Dyak   117   ccacatccccacctcactcctccgccacaatttcattccatttcagctcct~~~~~~~~c
Dana   117   gctc~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dsec   113   AGGTTCTTCCCGCCCCCTGGATCCACCATCCACCAGTGCCATGGAGGGAGTCG~~~~~~G
Dpse   121   aggttcttccc~~~~~ctggatccaccatccactagtgccatggagggagtca~~~~~~g
Dvir   47    cgtaattctgttttcaat~~~~~~~~tccatcaaagaaaataataaaactaatt~~~~tg
Dmoj   116   actgtcgcagttcagagctttataaatccatcaaaattaagaagagaagagaattctctg
Dgri   121   aatctacacataaaaaagttta~~~~~~~~~~~~~~~~~~~~~~~~~~~~caattctttg

                AFB
             --------
Dmel   153   TGACTCACTGGACGACAGTTT~~~~~GGCCAGCTTATTAAGTGCGCTACAAGCAGGCAAA
Dsim   153   TGACTCACTGGACGACAGTTT~~~~~GGCCAGCTTATTAAGTGCGATACAAGCTGGCAAA
Dsec   160   tgactcactggacgacagttt~~~~~ggccagcttattaagtgcgatacaagctggcaaa
Dere   149   tgactcagtggacgacagttt~~~~~ggccagcttattaagtgcgataaaa~~~cgcaaa
Dyak   169   tgactcagtggacgacagttt~~~~~ggccagcttattaagtgcgataaaa~~~ggcaaa
Dana   121   tgactcactgctggctcgatagcagaacttgaaaagaaaa~~ataccccttggcgggtta
Dsec   167   TGACTCACGAGAAAGAGAGAGACGGAGCCGTGGCTGGGGCTGGGGCTGGGGCTCTGATAA
Dpse   170   tgagtcacgagaaagagagagacggagccgtggctggggct~~~~~~~~~~~~ctgataa
Dvir   95    tgactcacgcgacgctctttgataaaaaaa~~~~accattaaaaa~gcgcagtagaaatg
Dmoj   176   tgactcatgcgtcgctcaacgagaagcaacaacaacaattaaaaaagcgcagtagaaatg
Dgri   153   tgagtcactcgcagtgataggcagcacttt~~~~~~~~~~~~aaaagcgcagtagcaata
```

```
Dmel   208   CA~~~AACAGCGCAGTTGGTGGAAATATCGAG~~~~~~~~~~~~~~~~~~~~~~~~~~~T
Dsim   208   CA~~~AACAGCGCAGTTGGTGGAAATATCGAG~~~~~~~~~~~~~~~~~~~~~~~~~~~T
Dsec   215   ca~~~aacagcgcagttggtggaaatatcgag~~~~~~~~~~~~~~~~~~~~~~~~~~~t
Dere   201   caacacacagcgcagttggtttaaaaatcgag~~~~~~~~~~~~~~~~~~~~~~~~~~~t
Dyak   221   ca~~~ttcagcgcagttggtggaaatatcgag~~~~~~~~~~~~~~~~~~~~~~~~~~~t
Dana   179   tgccatt~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~t
Dsec   227   AAAGGCAATTTACTCGTTAAAAGCGCAGTGAGAAGAGTGAAGGAAAAAAATGAGCTGAAT
Dpse   218   aaaggcaatttactcgttaaaagcgcagtgagaagagtgaaggaaaaaaatgagctgaat
Dvir   149   ~aaata~aaaaaaaacaaaa~~~~~ccgagaagttatcaaaaag~~~~~~ccggaagtgc
Dmoj   236   gaaa~aggagaagagaaaaaaaaacatgagtatttatcaataag~~~~~~ccggaagtgc
Dgri   201   aaacgc~~~~~~~~~~aaaaaaaacttaaatc~ttatcaaaaaaaaaaaaccggaagtgc

Dmel   238   TATCAATAAAACGGAGGTGCCCTAAACAAAACAAACACGAAAGCGAAAACAAAA~~~~~~
Dsim   238   TATCAATAAAACGGAAGTGCCCTAAACAAAACAAACACGAAAGCGAAAACAAAA~~~~~~
Dsec   245   tatcaataaaacggaagtgccctaaacaaaacaaacacgaaagcgaaaacaaaa~~~~~~
Dere   234   tatcaataaaacggaagtgccctaaacaaaacgaacatgaaagcgaaagcaaaa~~~~~~
Dyak   251   tatcaataaagcggaagtgccctaaacaaaacaaacacgaaagcgaaaatcaaaatgaaa
Dana   187   tatcaataaaacggaagtgccctaaacaaaacaaacatgaaagacgcgccaggcagc~~~
Dsec   287   TATCAATAAAACGGAAGTGAGAGAAGCAGAGGAGCACAGAGAGAGAGAGAGAGAGAGAGA
Dpse   278   tatcaataaaacggaagtgagagaagcagaggagcagagagagaga~~~~~~~~~~~~~~
Dvir   197   aaagcacagaacatacgagaaattagctgatatt~~gct~~~aataaagagagacatcac
Dmoj   289   aaagcgcagaacatacgagaaattagttgataaatcgcatggaataaaaaaaaaatacac
Dgri   250   aaagcacaaaacatgcgagaaattagttgataaatcacatgaaaagagacagatgaaggg

Dmel   292   CGGGCGAGACAGATAGT~CATAAATCAATGGAGCTTT~GCG~AAAAGATTCGCAGAAACG
Dsim   292   CGGGGGAGACAGATAGT~CATAAATCAATGGAGCTTG~GCG~AAAAGATTCCCAGAAACG
Dsec   299   cgggggagacagatagt~cataaatcaatggagcttt~gcg~aaaagattcccagaaacg
Dere   288   cggg~cagacagatagt~cataaatcaatggagcttt~gcg~taaagattcccagaaacg
Dyak   311   cgggggagacagatagt~cataaatcaatggagcttt~gca~aaaagattcccagaaacg
Dana   243   ~~~~~~~~~~~~~~~~~~cataaatcagaatatt~~~~~~~~~~~~~attcgcagaaacg
Dsec   347   AGGAAGGGATAGACAAAACATAAATTAAACTAAATAAAGCGCAAAAGATTCGCGGAAAAA
Dpse   324   aggaagggatagacaaaacataaattaaactaaataaagcgcaaaagattcgcggaaaaa
Dvir   252   tg~taa~agaga~~~~~~atgggtgtgagag~~~gtgggcggg~~~~~~gcgggggggag
Dmoj   349   tgctaatatagacacataaagagtgagagcgagtgtgggatgagtgatcagggggagggg
Dgri   310   a~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~agggggagtgt

Dmel   349   AAATTAAGTCATGCTAAAGATCGTCATTGACTGGAACTC~~~~~~~~~~~~~~~~AAAAT
Dsim   349   AAATTAAGTCATGCTAAAGATCGTCATTGACTGGAAATC~~~~~~~~~~~~~~~~CAAAT
Dsec   356   aaattaagtcatgctaaagatcgtcattgactggaaatc~~~~~~~~~~~~~~~~caaat
Dere   344   aaattaaatcatgctaaagatcgtcattgactggaactcggatcgaatggtgttcgaaat
Dyak   368   aaattaagtcatgctaaagatcgtcattgactggaactcgcatcgaatgctgttcgaaat
Dana   273   cattagagatcatcagatcagagactcgaag~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dsec   407   TGGCAGAATGGAATATCGTTT~GTCATTTAAGTGGAATTCCAAAACTCAGGCCGAAATAA
Dpse   384   tggcagaatggaatatcgttt~gtcatttaagtggaattccaaaactcaggtcgaaataa
Dvir   295   tacaa~~~~~~~ctggccgcatgcataaatcaaagtcgcatt~~~~aagcgaagcgaaac
Dmoj   409   tgtagggtattccagttcgc~tgcataaatcaaagtcatatt~~~~aagcgaagcgaaac
Dgri   322   cact~~~~~~~~~~~~~cgc~tgcataaataaatgaaaatcgtattcagcgaagcgaatc
```

```
Dmel   393  ACCCAAAAAAAGAAAAGAAAAAAAAACACCCTGGCCAAATTGAAATTATGATTAAGGCAA
Dsim   393  ACCCAAAAAA~GAAAAAA~~~~~~~~CACCCAGGCCAAATTGAAATTATGATTAAGGCAA
Dsec   400  acccaaaaaa~gaaaaaa~~~~~~~~cacccaggccaaattgaaattatgattaaggcaa
Dere   404  acccaaaaa~~gaaaaaa~~~~~~~~cacccagaccaaaatgaaattatgattaag~~~~
Dyak   428  acccaaaaa~~gaaaaaaa~~~~~~~cacccagaccaaaatgaaattatgattaagc~ag
Dana   304  acccaaacccaaaatcgagaaaaataaacagagccaggccggagccggagacac~~~~~~
Dsec   466  AATT~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~CAAAATAAAATTATGATTAAGCCAG
Dpse   443  aatt~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~caaaataaaattatgattaagccag
Dvir   344  acgaatgaattcgtagaaatgtca~~~acagtgacgtcacaatttcagaagaaa~cacac
Dmoj   463  ~~gaatgaattcgtagaaatgtcagtgacagtgacgtcacagtttcaacagaaaacagac
Dgri   367  ~~atatgaattcgtagaaatgtcacaat~~gtgacgtcacaatcttctaatt~~~~~~~~

Dmel   453  GCGGCAG~~~CAGGCCACCAAATGGCAAAAATTGGTCAACAC~~~AATTCTGTGACGTTT
Dsim   444  GCGG~~~~~~CAGCGGGCCAAATGGCAAAAATTGGTCAACAC~ATTGGTCTGTGACGTTT
Dsec   451  gcgg~~~~~~cagcaggccaaatgtcaaaaattggtcaaaac~attggtctgtgacgttt
Dere   449  ~~~~~~~~~~cagcaggccaaatggcaaaaattggtcaacacggc~~~tctgtgacgttt
Dyak   478  gcagcagg~~cagcaggccaaatggcaaaaattggtcaacacggc~~~tctgtgacgttt
Dana   357  ~~~~~~~~caggccaggccaggttgcaaaatttggccagcagctg~~~~~~~tgacgttt
Dsec   495  AAAAAGACACACACACAGACACACACACAG~~~~~~~~~~~~~~~~~~~CTGTGACGTTT
Dpse   472  aaaaagacacagacacacacacacaca~~g~~~~~~~~~~~~~~~~~~~ctgtgacgttt
Dvir   400  accaaaa~~~tgaattcaattaaaagctgttttaagaccgaaattccccacgctccgcca
Dmoj   522  cacagaaaactgaattaaattaaaagctgtttt~agacagaaattccccctgct~~~~~t
Dgri   415  ~~~~~~~~~~~~aattgacttaaaagccgttttaagcacaaaattccccactcactgcat

Dmel   507  GATAAGCTGTTGGGGCCAAAAAGGTGACTCTCGCCCCGAATCGCGGCTCAATAG~~~~~~
Dsim   497  GATAAGCTGCTGGGGCCAAAAAGGTGGCTCT~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dsec   504  gataagctgctggggccaaaaaggtggctct~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dere   497  gataagctgctggggccaaaaaggtggctctcgtctcgaattgcggctcaataa~~~~~~
Dyak   533  gataagctgctggggccaaaaaggtggctctcgcctcgaattgcggctcaataaccaata
Dana   403  gataagctgctgaggccaaaagtggctctaaagagcat~~~~~~~~~~~aatat~~~~~~
Dsec   536  GATGGGCTTCGTAT~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dpse   511  gatgggctccgtat~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dvir   457  gcgtttaagcaagttatgaataagcgtcaaaatgacgcgactaat~~~ttctgataagct
Dmoj   576  gcctttaagcaagttatgaataagcgtcaaagtgacgcgtctaag~~~ttctgataagc~
Dgri   464  catattaaccactttatgaataagcgttaaaatgacgacgcgactaatttctgataagc~

Dmel   560  ~~~~~~~~CCATAGCGAA~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dsim   527  ~~~~~~~~CCAAAGCGAA~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dsec   534  ~~~~~~~~ccaaagcgaa~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dere   550  ~~~~~~~~ccaaagcgaa~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dyak   593  acccataaccaaagcgaa~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dana   445  ~~~~~~~~ccatggcgaa~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dsec   549  ~~~~~~~~~~~~~~~~~~CTGTGGTTGCTGTATCTGTATCTGTATC~~~~~~~~~~~~~~
Dpse   524  ~~~~~~~~~~~~~~~~~~ctgtggttgctgtatctgtatctgtatc~~~~~~~~~~~~~~
Dvir   514  acgactaaaatg~~~~~~~~~~~~cgaatgcgtttgca~gcagagc~~~~~~~~~~gaga
Dmoj   631  ~~~~~~~~~~tg~~~~~~~~~~~~cgattgcgtttgca~ccagcaccagcacca~agaga
Dgri   522  ~~~~~~~~~~tgtttctaaaaatgcgcttgcgtttgcaaccagccaaagttggagagaga
```

```
Dmel   570  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dsim   537  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dsec   544  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dere   560  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dyak   610  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dana   455  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dsec   577  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dpse   552  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dvir   551  cag~~~~agacag~~~~~~~agagagagagagagagagaacat~~~~~~~~~~~~~~~~~
Dmoj   668  cac~~~~tgacagaaagaga~cggagagatggagaaagagagc~~~~~~~~~~~~~~~~~
Dgri   573  caacgagtgaaagaaagaaa~gaaagagagagagagagagagaggaagatagaggggg

                    GRH
                 --------
Dmel   570  ~TGTAAACCGGTTGGACAGTCAGTCAGTCAGTCAGTCAGTTAG~~~~~~~~~~TAAGTAA
Dsim   537  ~TGTAAACCGGTTGGACAGTCAGTCAGTCAGTTAG~~~~~~~~~~~~~~~~~~TAAGTAA
Dsec   544  ~tgtaaaccggttggacagtcagtcagtcagttag~~~~~~~~~~~~~~~~~~taagtaa
Dere   560  ~tgtaaaccggttggacagtcagtcagtcagtaag~~~~~~~~~~~~~~~~~~~~~~caa
Dyak   610  ~tgtaaaccggttggacagtcagtcagttagtaag~~~~~~~~~~~~~~~~~~~~~~caa
Dana   455  ~tggaaaccggttggccggtcactcagttagtcagt~~~~~~~~~~~~~~~~~~~~~caa
Dsec   577  ~TGTTAACCGGTTAACCATTCAGTCAATAA~~~~~~~~~~~~~~~~~~~~~~~~~~~~GT
Dpse   552  ~tgttaaccggttaaccattcagtcagtca~~~~~~~~~~~~~~~~~~~~~~~~ataagt
Dvir   582  ~tgaaaaccggtt~aactgtctgtcaataaaaccaaaaatccagt~~~~~~~~aaatgcc
Dmoj   705  ~tgcaaaccggtt~aactgcctgtcaataaaaccaaaaatccagt~~~~~~~~aaatgcc
Dgri   632  attaaaaccggtt~gactgacggtcaataaaaatcaaaatccagt~~~~~~~~aaatgcc

Dmel   620  TC~GGCGTAAAGTCGGCTAAAA~CCATAGCCAAATA~~~~~~AATACCAACGGAATGAGA
Dsim   579  TC~GGCGAAAAGTCGGCTAAAA~CCACAGCCAAATA~~~~~~AATACCAACGAAATGAGA
Dsec   586  tc~aacgaaaagtcggctaaaa~ccacagcccaata~~~~~~aataccaacgaaatgaga
Dere   598  tc~ggcgtaatgtcggctaaaa~ccactgccaaata~~~~~~aataccaacgaaatgaga
Dyak   648  tc~gacgtaaagtcggctaaaaaccactgccaaatactcgtaaataccaacgaaatgagg
Dana   494  tcaggcaagaagtcggctgaaaaccgcagccaaatacatg~~~~~~ccaacggaatgagg
Dsec   609  TGGCCAAAAGCACAGCCAAAGACACAGAAACAGCAGCA~~G~~~~~~ACGGCAGACAGCA
Dpse   588  tggccaaaagaacagccaaagacacagaaacagaaaca~~gcagcagacggcagacagca
Dvir   633  taagaaatgacacacaa~~gcacacacacacacacacacatacatgacacagcgca~~~~
Dmoj   756  taagaaatgacacacacatgctcacactcacacacacacacacgc~actgcgcacgca
Dgri   683  taagaaatgacacacacaggcagaca~~~~~~~~~~~~~~~~~~~~~~acagagcg~~ga

Dmel   672  CATACAGCGAGC~~~~~AAGTGGATGGACTCCGTCCC~ATCCGTTACTTTTTGAGTGCGT
Dsim   631  CATACAGCGAGC~~~~~AAGTGGATGGAC~~~~TCCGCATCCGTTACTTTTTAAGTGCGT
Dsec   638  catacagcgagc~~~~~aagtggatggac~~~~tccgcatccgttactttttaagtgcgt
Dere   650  catacaacgagc~~~~~aagtggatggac~~~~~gcccatccgttactttttaagtgcgt
Dyak   707  catacggcgagc~~~~~aagtggatggac~~~~ttcgcatccgttactttttaagtgcgt
Dana   548  cata~gcagagctgagcaagtggatggacttcagaagttctagtacaggta~~~~~~~~~
Dsec   661  GACAGCAAAATGCCAATGAAATGAGACTGAAGCAAGTGGCTGTCCCGTTCCCGGTTCTAA
Dpse   646  gacggcaaaatgccaatgaaatgagactgaagcaagtggctgtcccgttcccggttctag
Dvir   687  aaatgaacgaac~~~~~cgtaaccaacaagtggacaggcgagaaag~~~~~~~~~~~~~~
Dmoj   815  aaatgaacgaag~~~~~cgtagccaacaagtggacaacagagacag~~~~~~~~~~~~~~
Dgri   719  aaacgaacga~gcgaagcgtagccaacaagtgga~caacaacaacaacgatgaagaaaaa
```

```
Dmel   726  TCGAGTTCCTAGTCGTCACATGCAGATA~~~~~~CAGATACATAT~~~~~~~~~~~~~~~
Dsim   682  TCGAGTTCCTAG~~~CCACATGCAGATA~~~~~~CAGATACAGAT~~~~~~~~~~~~~~~
Dsec   689  tcgagttcctag~~~tcacatgcagata~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dere   700  ccgagttcctag~~~tcacatgcagatacagacacagatacagat~~~~~~~~~~~~~~~
Dyak   758  tcgagttcctag~~~tcacatgcagata~~~~~~cagatacagtt~~~~~~~~~~~~~~~
Dana   597  ~~~~~~~~~~~~~~~~~~~~~~cagata~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dsec   721  A~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dpse   706  a~~~~~~~~~~~~~gtcacac~~~~at~~~~~~gcagatacagat~~~~~~~~~~~~~~~
Dvir   727  ~~~~~~~~~~~~~~ag~~~acagcaaca~ctttgcccgttaatctcatagtgcgttcta~
Dmoj   855  ~~~~~~~~~~~~~~caataacaataacaacatacgccgttattcttgtagtgcgttcta~
Dgri   777  acgagaacctcaaccaataacaataacaaaatat~ccgttactcttgtagtgagttctaa

Dmel   764  ~~~~~~~~~~~~~~~ACAGATACAGAAACACACAATCAGAATCAGATACA~~~~~~~~~~
Dsim   717  ~~~~~~~~~~~~~~~~~~~~~~~ACAAACACACAATCAGAATCAGATACA~~~~~~~~~~
Dsec   713  ~~~~~~~~~~~~~~~~~~~~~~~~aaaacacacaatcagaatcagataca~~~~~~~~~~
Dere   741  ~~~~~~~~~~~~~~~~~acatatacaactacacaatcagaatcagataca~~~~~~~~~~
Dyak   793  ~~~~~~~~~~~~~~~~~acagatacaaatacacaatgagaatcagataca~~~~~~~~~~
Dana   603  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~cagatacagataca~~~~~~~~~~
Dsec   721  ~~~~~~~~~~~~~~~~~~~~~~GTCACAC~~~~AT~~~~~~GCAGATACAGATACA~~~~
Dpse   727  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~acagataca~~~~~~~~~~
Dvir   768  ~~~~~~~ccgctggcgaca~~~~~~~~~~~~~~tgcacga~~aaggaggaatgggaa~~~
Dmoj   900  ~~~~~~~~~~~~~gcgacaagtcaca~~~~~~~tgcatgc~~aagaagcagactg~a~~~
Dgri   836  cgctggctgattggcgacaagtcacaagttccatgcatgcgaaagaaagaaagacactcg

Dmel   800  CAAAGTATCTG~~~~~~~~~~GGGGCATTACTCATGCTAATTT~~~~~~~~~~~~~~~~~
Dsim   745  CAAAGTATCTG~~~~~~~~~~GCGGCATTACTCATGCTAATTT~~~~~~~~~~~~~~~~~
Dsec   740  caaagtatctg~~~~~~~~~~gcggcattactcatgctaattt~~~~~~~~~~~~~~~~~
Dere   775  caaagtatctg~~~~~~~~~~ggggcattactcatgctaattt~~~~~~~~~~~~~~~~~
Dyak   827  caaagtatctg~~~~~~~~~~ggggcattactcatgctaattt~~~~~~~~~~~~~~~~~
Dana   618  tagagtatct~~~~~~~~~~~~aggcattactcatgctaataa~~~~~~~~~~~~~~~~~
Dsec   746  GAAAGTATCTGCTGCTT~~~CTGGGCATTACTCATGCTAATTT~~~~~~~~~~~~~~~~~
Dpse   737  gaaagtatctgctgctgctgctgggcattactcatgctaattt~~~~~~~~~~~~~~~~~
Dvir   803  aaaagcatct~~~acgc~~~~~~~gtattactcatgctaattttacacacacacacacac
Dmoj   935  ataagcatctcaacggc~~~~~~~gtattactcatgctaattccacacagacaca~~~~~
Dgri   896  aaaagcatctctgccttc~~~~~~ttattactcatgctaattccgtacacaca~~~~~~~

Dmel   832  ~~~~~~~~~~~~~CACACAGATGCTCGGCCACTGCGGA~~~~~~~~~~~~~~~~~~~~~~
Dsim   777  ~~~~~~~~~~~~~CACACAGATGCTCGGCCACTGCGGA~~~~~~~~~~~~~~~~~~~~~~
Dsec   772  ~~~~~~~~~~~~~cacacagatgctcggccactgcgga~~~~~~~~~~~~~~~~~~~~~~
Dere   807  ~~~~~~~~~~~~~cacacagatgctcggccactgcgga~~~~~~~~~~~~~~~~~~~~~~
Dyak   859  ~~~~~~~~~~~~~cacacagatgcttggccactgcgga~~~~~~~~~~~~~~~~~~~~~~
Dana   648  ~~~~~~~~~~~~~tcactacactgagagaagcatattt~~~~~~~~~~~~~~~~~~~~~~
Dsec   785  ~~~~~~~~~~~~~CACACTCACACAGAGACAGATACAG~~~~~~~~~~~~~~~~~AGACA
Dpse   779  ~~~~~~~~~~~~~cacactcacacagatacagatacag~~~~~~~~~~~~~~~~~ataca
Dvir   853  acacgcgcacacacacacacatgctcggctgga~~~~~~~gtcag~~~~~~~~~~tgttg
Dmoj   982  ~~~~~~~~~~~~~cacacaca~~~~~~~~~~~~~~~~~~~gccag~~~~~~~~~~tgttg
Dgri   942  ~~~~~~~~~~~~~cacacacactcaatcaggcagtgagagtccag~~~~~~~~~~tgttg
```

```
Dmel    857  ~~~~~~~~~~~~~~~~~ATAATGAACAACTTAAAGCGCCAAAGTCGTCGCCGAGTTGAGT
Dsim    802  ~~~~~~~~~~~~~~~~~ATAATGAACAACTTAAAGCGCCAAAGACGTCGCCGAGTTGAGT
Dsec    797  ~~~~~~~~~~~~~~~~~ataatgaacaacttaaagcgccaaagacgtcgccgagttgagt
Dere    832  ~~~~~~~~~~~~~~~~~ataatgaactacttaaagcgcgaaagacgtcgccgagttgagt
Dyak    884  ~~~~~~~~~~~~~~~~~ataatgaactacttaaagcgccaaagacgtcgccggattgagt
Dana    673  ~~~~~~~~~~~~~~~~~cctgtgaaagtatctttttgagcggggagtgggtgttggcag~
Dsec    816  GAGACAGCGACAACAGTATAATGAACTACTTTAAGGCGTCCAGAG~~~~CCGAGTTGCGT
Dpse    810  gagacagcgacaacagtataatgaactactttaaggcgtccagag~~~~ccgagttgcgt
Dvir    896  gacagt~~~~~~~~~~~ataatgaactacttaatgccgcacagcacagtagtggcagcac
Dmoj    1001 ggcagtaagt~~~~~~~ataatgaactacttaa~~~~~~~~~~~~~~~~~~~~~~~~ctc
Dgri    980  gacagt~~~~~~~~~~~ataatgaactacttaatgccgcagcagcagtg~~~~~~~~~~~

                  GRH
                --------
Dmel    901  T~~AACAAGTTCACAAAGAACTGCGGGTACACAGCAAACAAAACTTGCGCCAAATTTTAT
Dsim    846  T~~AACAAGTTCTCAAAGAACTGCGGGTACACAGTAAACAAAACTTGTGGCTCATACAAT
Dsec    841  t~~aacaagttctcaaagaactgcggatacacagtaaacaaaacttgtgcctcatacaat
Dere    876  t~~aacaagttttcaaagaactgcgggtacacagtaaagaagatttgtgccttgtttaat
Dyak    928  t~~aacaagttttcaaagaactgcgactacacagtaaacaagatttg~gccttattcctt
Dana    715  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dsec    872  T~~AACAAGTTTTCTTGAGCAGCCGCACAGTGGGCTGGCCTCCCATGGTTACCACGTGCC
Dpse    866  t~~aacaagttttcttgagcagccgcacagtgggctggcctcccatggttaccacgtgcc
Dvir    945  ggaaacaagttgtcg~~~cacagtgggcaaagtgtgtcttatactgtgcact~~~~~~~~
Dmoj    1030 agaaacaagttgtcg~~~caccgtgggcaaagagatacttattctgtgtactctatatac
Dgri    1017 ~~aaacaagttgtcgtcgcacagtgggcaaagttg~~cttatgtattgta~tgtatgtat

                                                                  GRH
                                                                --------
Dmel    959  ATTCGATTCAAAGAAATATTCTTAATATTTTA~~~~~~~~TTTATTCATGGCAACTTGTT
Dsim    904  TTTCGAT~~~~TATTATAATTATTATATTTGA~~~~~~~~TTGGTTCATGGCAACTTTTT
Dsec    899  tttcgat~~~~tattattattattatatttga~~~~~~~~ttggttcatggcaacttttt
Dere    934  tttccatacgcagaaaaatacatttcaatggtattgattggttgttcacggacacatttt
Dyak    985  tttaaattcgcaggaaaataca~~~~~~~~~ttcttatggtatcaatttatagtattgta
Dana    715  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dsec    930  ACTTGCCACTTGCCACAGAGCAGAGCCCTAGATCGGGGCAGAGACTGGTACCTGGTACCT
Dpse    924  acttgccacttgccacagagcagagccctagatcggggcagagactggtgcctggtacct
Dvir    993  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~tacatatattgtaatcgcattgtt~~~~~
Dmoj    1087 atacatatttgcatacacacatatatatgaatatgtatattgtaatcgcattgttgttg~
Dgri    1073 atatatatttatatatatatatatatatatgtatttcaattgcatta~~~~~ttgttgttgt

Dmel    1011 CAAAAGTGTTTTACAATATAATCTGCAAAACCTTAACCACTTCAGTTGTGTGGGAGGTTT
Dsim    952  CCCAAGTGTCTTACATTTTGATCTGCGTAACCCTAACAACTTCAGTTGTGTGGGAGATTT
Dsec    947  cccaagtgtcttacattttaatctgcgtaaccctaacaacttcagttgtgtgggagattt
Dere    994  gccaagtgttttacgatttgatctgcggaaccctaaccacttcagttgtgtgggagattt
Dyak    1036 tctgattggtttatgag~~~atctgcgaaaccctaacaacttgagttgtgtgggagatta
Dana    715  ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dsec    990  GGCACCTGGTACCTGTAGGCCTGTGTGGAGTGTGGAGTGTGGGTCGCAAAAACCCATTGA
Dpse    984  g~~~~~~~~~~~~~~~~~~~~~~~taggcctgtggagtgtgggtcgcaaaaacccattga
Dvir    1017 ~~~~~~~~~~~~~~~~~~~~~accacgtgccacttgcagctttgcgtggtaa~~caccac
Dmoj    1145 ~~~~~~~~~~~~~~~~ttgttaccacgtgccacttgcagcttagcgtggtaa~~caccac
Dgri    1128 tgcttaagtattttttttgttaccacgtgccacttgcagcttagcgtggtaacccaacac
```

```
Dmel   1071 CAATTGACTGGCATTCCTTTGGCAAAATTGTAGGCCATAAAGATATGAAATTGCAGAGAC
Dsim   1012 CAGTTGACTGGCATTTCTTTGGCCATATTTTGGGCCATAAAGATATGAAATTGCAGAGAC
Dsec   1007 caattgactggcatttctttggccatattttgagccataaagatatgaaatttcagagac
Dere   1054 caattgagtggcatttctttggcaaaatttcaggccataaacatatgaaattacagagat
Dyak   1093 caattgactggcatttcttgggcaaacttcgaggccataaacatatgaaattacagagat
Dana   715  ~~~~tgactggcagctctagccc~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Dsec   1050 AAGACCTATGAAATTTAATAGATTTCATGAAAGTATCCATCGGATT~~~~~~~~~~~~~~
Dpse   1021 aagacctatgaaatttaatagatttcatgaaagtatccatcggatt~~~~~~~~~~~~~~
Dvir   1055 aaatgaggtcgtcggcgttatgaagcaggaggcgtgtagacgtgggctga~~~~~~~~~~
Dmoj   1188 aaatgaggtcgccgcta~~~~~~~~~aggaggcgtgtagacgtgggccgcagaaaaaaaa
Dgri   1188 aaatgaggtcgtcagtgttccgaaggagactcaac~~agacgtgggccgtcaatatgact

Dmel   1131 TTTT~GAAAGCTGCCAATGCCAACTGATTGACAGCAGGATCCTT
Dsim   1072 TTTT~GAAAGCTGCCAATGCCAACTGATTGACAGCAGAATTTCG
Dsec   1067 tttt~gaaagctgccaatgccaactgattgacagcagaatttcg
Dere   1114 tttt~gaaagctgccagtgccaattgattgacagcagaatttcg
Dyak   1153 ttttcgaaagctgccaatgccaactgattgacagctgaatttcg
Dana   734  ~~~~~~~~~~~~~~~~~~~~~~~~~gattgacagcgc
Dsec   1095 ~~~~~~~~~~~~~~~~~~~~~~~~GGATTGACAG
Dpse   1066 ~~~~~~~~~~~~~~~~~~~~~~~~ggattgacag
Dvir   1104 ~~~~~~~~~~~ctttttgcgtattggccgcatttctaagccgtt
Dmoj   1239 tctacttttttttttttgcaaactggccgcatttcttagccgaa
Dgri   1246 ata~~~~~ttttttttagcgaattggccgcatttctaagccgaa
```

References

Alibardi, L. and Kwang, W. J. (2006). Structural and Immunocytochemical Characterization of Keratinization in Vertebrate Epidermis and Epidermal Derivatives. In *International Review of Cytology*, vol. Volume 253, pp. 177-259: Academic Press.

Andersen, S. O., Hojrup, P. and Roepstorff, P. (1995). Insect cuticular proteins. *Insect Biochemistry and Molecular Biology* **25**, 153-176.

Andrew, D. J., Horner, M. A., Petitt, M. G., Smolik, S. M. and Scott, M. P. (1994). Setting limits on homeotic gene function: restraint of Sex combs reduced activity by teashirt and other homeotic genes. *EMBO J.* **13**, 1132-1144.

Andrioli, L. P., Vasisht, V., Theodosopoulou, E., Oberstein, A. and Small, S. (2002). Anterior repression of a Drosophila stripe enhancer requires three position-specific mechanisms. *Development* **129**, 4931-40.

Appel, B. and Sakonju, S. (1993). Cell-type-specific mechanisms of transcriptional repression by the homeotic gene products UBX and ABD-A in Drosophila embryos. *EMBO J.* **12**, 1099-1109.

Arenas-Mena, C., Cameron, A. R. and Davidson, E. H. (2000). Spatial expression of Hox cluster genes in the ontogeny of a sea urchin. *Development* **127**, 4631-4643.

Arenas-Mena, C., Martinez, P., Cameron, R. A. and Davidson, E. H. (1998). Expression of the Hox gene complex in the indirect development of a sea urchin. *Proc. Natl Acad. Sci. USA* **95**, 13062-13067.

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

Azpiazu, N. and Morata, G. (1998). Functional and regulatory interactions between Hox and extradenticle genes. *Genes Dev.* **12**, 261-273.

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

Barolo, S., Castro, B. and Posakony, J. W. (2004). New Drosophila transgenic reporters: insulated P-element vectors expressing fast-maturing RFP. *Biotechniques* **36**, 436-40, 442.

Beeman, R. W., Stuart, J. J., Haas, M. S. and Denell, R. E. (1989). Genetic analysis of the homeotic gene complex (HOM-C) in the beetle Tribolium castaneum. *Dev. Biol.* **133**, 196-209.

Bello, B. C., Hirth, F. and Gould, A. P. (2003). A pulse of the Drosophila Hox protein Abdominal-Aschedules the end of neural proliferation via neuroblast apoptosis. *Neuron* **37**, 209-219.

Benbrook, D. M. and Jones, N. C. (1994). Different binding specificities and transactivation of variant CRE's by CREB complexes. *Nucleic Acids Res* **22**, 1463-9.

Berman, B. P., Nibu, Y., Pfeiffer, B. D., Tomancak, P., Celniker, S. E., Levine, M., Rubin, G. M. and Eisen, M. B. (2002). Exploiting transcription factor binding site clustering to identify cis-regulatory modules involved in pattern formation in the Drosophila genome. *Proc Natl Acad Sci U S A* **99**, 757-62.

Bienz, M. (1994). Homeotic genes and positional signalling in the Drosophila viscera. *Trends Genet.* **10**, 22-26.

Bier, K. and Müller, W. (1969). DNA-Messungen bei Insekten und eine Hypothese über retardierte Evolution und besonderen DNA-Reichtum in Tierreich. *Biologisches Zentralblatt* **88**, 425-449.

Boffelli, D., McAuliffe, J., Ovcharenko, D., Lewis, K. D., Ovcharenko, I., Pachter, L. and Rubin, E. M. (2003). Phylogenetic Shadowing of Primate Sequences to Find Functional Regions of the Human Genome. *Science* **299**, 1391-1394.

Bokoch, G. M. (2005). Regulation of innate immunity by Rho GTPases. *Trends Cell Biol* **15**, 163-71.

Bonneton, F., Shaw, P. J., Fazakerley, C., Shi, M. and Dover, G. A. (1997). Comparison of bicoid-dependent regulation of hunchback between Musca domestica and Drosophila melanogaster. *Mech Dev* **66**, 143-56.

Bray, S. J. and Hirsh, J. (1986). The Drosophila virilis dopa decarboxylase gene is developmentally regulated when integrated into Drosophila melanogaster. *Embo J* **5**, 2305-11.

Bray, S. J., Johnson, W. A., Hirsh, J., Heberlein, U. and Tjian, R. (1988). A cis-acting element and associated binding factor required for CNS expression of the Drosophila melanogaster dopa decarboxylase gene. *Embo J* **7**, 177-88.

Brenowitz, M., Senear, D. F., Shea, M. A. and Ackers, G. K. (1986). Quantitative DNase footprint titration: a method for studying protein-DNA interactions. *Methods Enzymol* **130**, 132-81.

Brock, J., Midwinter, K., Lewis, J. and Martin, P. (1996). Healing of incisional wounds in the embryonic chick wing bud: characterization of the actin purse-string and demonstration of a requirement for Rho activation. *J Cell Biol* **135**, 1097-107.

Bromleigh, V. C. and Freedman, L. P. (2000). p21 is a transcriptional target of HOXA10 in differentiating myelomonocytic cells. *Genes Dev.* **14**, 2581-2586.

Bruhl, T. (2004). Homeobox A9 transcriptionally regulates the EphB4 receptor to modulate endothelial cell migration and tube formation. *Circ. Res.* **94**, 743-751.

Capovilla, M. and Botas, J. (1998). Functional dominance among Hox genes: repression dominates activation in the regulation of dpp. *Development* **125**, 4949-4957.

Capovilla, M., Brandt, M. and Botas, J. (1994). Direct regulation of decapentaplegic by Ultrabithorax and its role in Drosophila midgut morphogenesis. *Cell* **76**, 461-475.

Capovilla, M., Kambris, Z. and Botas, J. (2001). Direct regulation of the muscle-identity gene apterous by a Hox protein in the somatic mesoderm. *Development* **128**, 1221-1230.

Castelli-Gair, J. and Akam, M. (1995). How the Hox gene Ultrabithorax specifies two different segments: the significance of spatial and temporal regulation within metameres. *Development* **121**, 2973-82.

Chan, S. K. (1997). Switching the in vivo specificity of a minimal Hox-responsive element. *Development* **124**, 2007-2014.

Chan, S. K., Popperl, H., Krumlauf, R. and Mann, R. S. (1996a). An extradenticle-induced conformational change in a HOX protein overcomes an inhibitory function of the conserved hexapeptide motif. *EMBO J.* **15**, 2476-2487.

Chan, S. K., Ryoo, H. D., Gould, A., Krumlauf, R. and Mann, R. S. (1996b). Switching the in vivo specificity of a minimal Hox-responsive element. *Development* **124**, 2007-2014.

Chang, C. P. (1995). Pbx proteins display hexapeptide-dependent cooperative DNA binding with a subset of Hox proteins. *Genes Dev.* **9**, 663-674.

Chauvet, S. (2000). dlarp, a new candidate Hox target in Drosophila whose orthologue in mouse is expressed at sites of epithelium/mesenchymal interactions. *Dev. Dyn.* **218**, 401-413.

Chen, J. and Ruley, H. E. (1998). An enhancer element in the EphA2 (Eck) gene sufficient for rhombomere-specific expression is activated by HOXA1 and HOXB1 homeobox proteins. *J. Biol. Chem.* **273**, 24670-24675.

Clark, S. G., Chisholm, A. D. and Horvitz, H. R. (1993). Control of cell fates in the central body region of C. elegans by the homeobox gene lin-39. *Cell* **74**, 43-55.

Cobb, J. and Duboule, D. (2005). Comparative analysis of genes downstream of the Hoxd cluster in developing digits and external genitalia. *Development* **132**, 3055-3067.

Crooks, G. E., Hon, G., Chandonia, J. M. and Brenner, S. E. (2004). WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188-1190.

Cui, M. and Han, M. (2003). Cis regulatory requirements for vulval cell-specific expression of the Caenorhabditis elegans fibroblast growth factor gene egl-17. *Dev. Biol.* **257**, 104-116.

de Zulueta, P., Alexandre, E., Jacq, B. and Kerridge, S. (1994). Homeotic complex and teashirt genes co-operate to establish trunk segmental identities in Drosophila. *Development* **120**, 2287-2296.

Ebner, A., Cabernard, C., Affolter, M. and Merabet, S. (2005). Recognition of distinct target sites by a unique Labial/Extradenticle/Homothorax complex. *Development* **132**, 1591-1600.

Ekker, S. C., Young, K. E., von Kessler, D. P. and Beachy, P. A. (1991). Optimal DNA sequence recognition by the Ultrabithorax homeodomain of Drosophila. *Embo J* **10**, 1179-86.

Eresh, S., Riese, J., Jackson, D. B., Bohmann, D. and Bienz, M. (1997). A CREB-binding site as a target for decapentaplegic signalling during Drosophila endoderm induction. *Embo J* **16**, 2014-22.

Erives, A. and Levine, M. (2004). Coordinate enhancers share common organizational features in the Drosophila genome. *Proc Natl Acad Sci U S A* **101**, 3851-6.

Estella, C., Rieckhof, G., Calleja, M. and Morata, G. (2003). The role of buttonhead and Sp1 in the development of the ventral imaginal discs of Drosophila. *Development* **130**, 5929-41.

Fasano, L. (1991). The gene teashirt is required for the development of Drosophila embryonic trunk segments and encodes a protein with widely spaced zinc finger motifs. *Cell* **64**, 63-79.

Finnerty, J. R., Pang, K., Burton, P., Paulson, D. and Martindale, M. Q. (2004). Origins of bilateral symmetry: Hox and dpp expression in a sea anemone. *Science* **304**, 1335-1337.

Fried, M. and Crothers, D. M. (1981). Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide gel electrophoresis. *Nucleic Acids Res* **9**, 6505-25.

Galant, R., Walsh, C. M. and Carroll, S. B. (2002). Hox repression of a target gene: extradenticle-independent, additive action through multiple monomer binding sites. *Development* **129**, 3115-3126.

Galko, M. J. and Krasnow, M. A. (2004). Cellular and genetic analysis of wound healing in Drosophila larvae. *PLoS Biol* **2**, E239.

Gao, J. and Scott, J. G. (2006). Use of quantitative real-time polymerase chain reaction to estimate the size of the house-fly Musca domestica genome. *Insect Mol Biol* **15**, 835-7.

Garcia-Bellido, A. (1977). Homeotic and atavic mutations in insects. *Am. Zool.* **17**, 613-629.

Gebelein, B., Culi, J., Ryoo, H. D., Zhang, W. and Mann, R. S. (2002). Specificity of Distalless repression and limb primordia development by abdominal Hox proteins. *Dev. Cell* **3**, 487-498.

Gebelein, B., McKay, D. J. and Mann, R. S. (2004). Direct integration of Hox and segmentation gene inputs during Drosophila development. *Nature* **431**, 653-659.

Gibert, J. M. and Simpson, P. (2003). Evolution of cis-regulation of the proneural genes. *Int J Dev Biol* **47**, 643-51.

Giesen, K., Lammel, U., Langehans, D., Krukkert, K., Bunse, I. and Klambt, C. (2003). Regulation of glial cell number and differentiation by ecdysone and Fos signaling. *Mech Dev* **120**, 401-13.

Goto, S. and Hayashi, S. (1997). Specification of the embryonic limb primordium by graded activity of Decapentaplegic. *Development* **124**, 125-32.

Gould, A., Morrison, A., Sproat, G., White, R. A. H. and Krumlauf, R. (1997). Positive cross-regulation and enhancer sharing: two mechanisms for specifying overlapping Hox expression patterns. *Genes Dev.* **11**, 900-913.

Gould, A. P. and White, R. A. (1992). Connectin, a target of homeotic gene control in Drosophila. *Development* **116**, 1163-1174.

Graba, Y. (1992). Homeotic control in Drosophila; the scabrous gene is an in vivo target of Ultrabithorax proteins. *EMBO J.* **11**, 3375-3384.

Graba, Y. (1995). DWnt-4, a novel Drosophila Wnt gene acts downstream of homeotic complex genes in the visceral mesoderm. *Development* **121**, 209-218.

Grieder, N. C., Marty, T., Ryoo, H. D., Mann, R. S. and Affolter, M. (1997). Synergistic activation of a Drosophila enhancer by HOM/EXD and DPP signaling. *EMBO J.* **16**, 7402-7410.

Grienenberger, A. (2003). TGF-[beta] signaling acts on a Hox response element to confer specificity and diversity to Hox protein function. *Development* **130**, 5445-5455.

Grossman, G. L., Rafferty, C. S., Clayton, J. R., Stevens, T. K., Mukabayire, O. and Benedict, M. Q. (2001). Germline transformation of the malaria vector, Anopheles gambiae, with the piggyBac transposable element. *Insect Mol Biol* **10**, 597-604.

GuhaThakurta, D. (2006). Computational identification of transcriptional regulatory elements in DNA sequence. *Nucleic Acids Res* **34**, 3585-98.

Haerry, T. and Gehring, W. (1997). A conserved cluster of homeodomain binding sites in the mouse Hoxa-4 intron functions in Drosophila embryos as an enhancer that is directly regulated by Ultrabithorax. *Dev. Biol.* **186**, 1-15.

Handler, A. M. and Harrell, R. A., 2nd. (1999). Germline transformation of Drosophila melanogaster with the piggyBac transposon vector. *Insect Mol Biol* **8**, 449-57.

Henderson, K. D. and Andrew, D. J. (2000). Regulation and function of Scr, exd, and hth in the Drosophila salivary gland. *Developmental Biology* **217**, 362-374.

Hersh, B. M. and Carroll, S. B. (2005). Direct regulation of knot gene expression by Ultrabithorax and the evolution of cis-regulatory elements in Drosophila. *Development* **132**, 1567-1577.

Heuer, J. G., Li, K. and Kaufman, T. C. (1995). The Drosophila homeotic target gene centrosomin (cnn) encodes a novel centrosomal protein with leucine zippers and maps to a genomic region required for midgut morphogenesis. *Development* **121**, 3861-3876.

Hoffmann, J. A. and Reichhart, J. M. (2002). Drosophila innate immunity: an evolutionary perspective. *Nat Immunol* **3**, 121-6.

Hooiveld, M. H. (1999). Novel interactions between vertebrate Hox genes. *Int. J. Dev. Biol.* **43**, 665-674.

Houghton, L. and Rosenthal, N. (1999). Regulation of a muscle-specific transgene by persistent expression of Hox genes in postnatal murine limb muscle. *Dev. Dyn.* **216**, 385-397.

I.H.G.S.C. (2004). Finishing the euchromatic sequence of the human genome. *Nature* **431**, 931-945.

Ishii, M. (1999). Hbox1 and Hbox7 are involved in pattern formation in sea urchin embryos. *Dev. Growth Differ.* **41**, 241-252.

Jeong, S., Rokas, A. and Carroll, S. B. (2006). Regulation of body pigmentation by the Abdominal-B Hox protein and its gain and loss in Drosophila evolution. *Cell* **125**, 1387-99.

Jessen, B. A., Qin, Q. and Rice, R. H. (2000). Functional AP1 and CRE response elements in the human keratinocyte transglutaminase promoter mediating Whn suppression. *Gene* **254**, 77-85.

Jiravanichpaisal, P., Lee, B. L. and Soderhall, K. (2006). Cell-mediated immunity in arthropods: hematopoiesis, coagulation, melanization and opsonization. *Immunobiology* **211**, 213-36.

Jones, N. C. and Pevzner, P. A. (2004). An Introduction to Bioinformatics Algorithms. Cambridge: MIT Press.

Karlsson, C., Korayem, A. M., Scherfer, C., Loseva, O., Dushay, M. S. and Theopold, U. (2004). Proteomic analysis of the Drosophila larval hemolymph clot. *J Biol Chem* **279**, 52033-41.

Kaufman, T. C., Seeger, M. A. and Olsen, G. (1990). Molecular and genetic organization of the antennapedia gene complex of Drosophila melanogaster. *Adv. Genet.* **27**, 309-362.

Kim, J. (2001). Macro-evolution of the hairy enhancer in Drosophila species. *J Exp Zool* **291**, 175-85.

Koh, K. (2002). Cell fates and fusion in the C. elegans vulval primordium are regulated by the EGL-18 and ELT-6 GATA factors [mdash] apparent direct targets of the LIN-39 Hox protein. *Development* **129**, 5171-5180.

Kosman, D. (2004). Multiplex detection of RNA expression in Drosophila embryos. *Science* **305**, 846.

Kosman, D., Mizutani, C. M., Lemons, D., Cox, W. G., McGinnis, W. and Bier, E. (2004). Multiplex Detection of RNA Expression in Drosophila Embryos. *Science* **305**, 846-.

Kremser, T. (1999). Expression of the [beta]3 tubulin gene ([beta]Tub60D) in the visceral mesoderm of Drosophila is dependent on a complex enhancer that binds Tinman and UBX. *Dev. Biol.* **216**, 327-339.

Krumlauf, R. (1994). Hox genes in vertebrate development. *Cell* **78**, 191-201.

Kuziora, M. A. and McGinnis, W. (1988). Autoregulation of a Drosophila homeotic selector gene. *Cell* **55**, 477-485.

Lampe, X., Picard, J. J. and Rezsohazy, R. (2004). The Hoxa2 enhancer 2 contains a critical Hoxa2 responsive regulatory element. *Biochem. Biophys. Res. Commun.* **316**, 898-902.

Lander, E. S. Linton, L. M. Birren, B. Nusbaum, C. Zody, M. C. Baldwin, J. Devon, K. Dewar, K. Doyle, M. FitzHugh, W. et al. (2001). Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921.

Lei, H., Wang, H., Juan, A. H. and Ruddle, F. H. (2005). The identification of Hoxc8 target genes. *Proc. Natl Acad. Sci. USA* **102**, 2420-2424.

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

Liu, J. and Fire, A. (2000). Overlapping roles of two Hox genes and the exd ortholog ceh-20 in diversification of the C. elegans postembryonic mesoderm. *Development* **127**, 5179-5190.

Lohmann, I., McGinnis, N., Bodmer, M. and McGinnis, W. (2002). The Drosophila Hox gene Deformed sculpts head morphology via direct regulation of the apoptosis activator reaper. *Cell* **110**, 457-466.

Lou, L., Bergson, C. and McGinnis, W. (1995). Deformed expression in the Drosophila central nervous system is controlled by an autoactivated intronic enhancer. *Nucleic Acids Res.* **23**, 3481-3487.

Ludwig, M. Z., Patel, N. H. and Kreitman, M. (1998). Functional analysis of eve stripe 2 enhancer evolution in Drosophila: rules governing conservation and change. *Development* **125**, 949-58.

Mace, K. A., Pearson, J. C. and McGinnis, W. (2005). An epidermal barrier wound repair pathway in Drosophila is mediated by grainy head. *Science* **308**, 381-5.

Maconochie, M. K. (1997). Cross-regulation in the mouse HoxB complex: the expression of Hoxb2 in rhombomere 4 is regulated by Hoxb1. *Genes Dev.* **11**, 1885-1895.

Magli, M. C., Largman, C. and Lawrence, H. J. (1997). Effects of HOX homeobox genes in blood cell differentiation. *J. Cell. Physiol.* **173**, 168-177.

Mahaffey, J. P., Griswold, C. M. and Cao, Q. (2001). The Drosophila genes disconnected and disco-related are redundant with respect to larval head development and accumulation of mRNAs from Deformed target genes. *Genetics* **157**, 225-236.

Manak, J. R., Mathies, L. D. and Scott, M. P. (1994). Regulation of a decapentaplegic midgut enhancer by homeotic proteins. *Development* **120**, 3605-3612.

Mann, R. and Affolter, M. (1998). Hox proteins meet more partners. *Curr. Opin. Genet. Dev.* **8**, 423-429.

Mann, R. S. (1994). Engrailed-mediated repression of Ultrabithorax is necessary for the parasegment 6 identity in Drosophila. *Development* **120**, 3205-12.

Mann, R. S. and Chan, S. K. (1996). Extra specificity from extradenticle: the partnership between HOX and PBX/EXD homeodomain proteins. *Trends Genet.* **12**, 258-262.

Markstein, M., Markstein, P., Markstein, V. and Levine, M. S. (2002). Genome-wide analysis of clustered Dorsal binding sites identifies putative target genes in the Drosophila embryo. *Proc Natl Acad Sci U S A* **99**, 763-8.

Markstein, M., Zinzen, R., Markstein, P., Yee, K. P., Erives, A., Stathopoulos, A. and Levine, M. (2004). A regulatory code for neurogenic gene expression in the Drosophila embryo. *Development* **131**, 2387-94.

Martin, P. and Lewis, J. (1992). Actin cables and epidermal movement in embryonic wound healing. *Nature* **360**, 179-83.

Martin, P. and Parkhurst, S. M. (2004). Parallels between tissue repair and embryo morphogenesis. *Development* **131**, 3021-34.

Masquilier, D. and Sassone-Corsi, P. (1992). Transcriptional cross-talk: nuclear factors CREM and CREB bind to AP-1 sites and inhibit activation by Jun. *J Biol Chem* **267**, 22460-6.

Mastick, G. S., McKay, R., Oligino, T., Donovan, K. and Lopez, A. J. (1995). Identification of target genes regulated by homeotic proteins in Drosophila melanogaster through genetic selection of Ultrabithorax protein-binding sites in yeast. *Genetics* **139**, 349-363.

Matsuo, I. and Yasuda, K. (1992). The cooperative interaction between two motifs of an enhancer element of the chicken alpha A-crystallin gene, alpha CE1 and alpha CE2, confers lens-specific expression. *Nucleic Acids Res* **20**, 3701-12.

McCluskey, J. and Martin, P. (1995). Analysis of the tissue movements of embryonic wound healing--DiI studies in the limb bud stage mouse embryo. *Dev Biol* **170**, 102-14.

McCormick, A., Core, N., Kerridge, S. and Scott, M. P. (1995). Homeotic response elements are tightly linked to tissue-specific elements in a transcriptional enhancer of the teashirt gene. *Development* **121**, 2799-2812.

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

Mehic, D., Bakiri, L., Ghannadan, M., Wagner, E. F. and Tschachler, E. (2005). Fos and jun proteins are specifically expressed during differentiation of human keratinocytes. *J Invest Dermatol* **124**, 212-20.

Merzendorfer, H. and Zimoch, L. (2003). Chitin metabolism in insects: structure, function and regulation of chitin synthases and chitinases. *J Exp Biol* **206**, 4393-412.

Mirny, L. A. and Gelfand, M. S. (2002). Structural analysis of conserved base pairs in protein-DNA complexes. *Nucleic Acids Res* **30**, 1704-11.

Morgan, R., Nalliah, A. and Morsi El-Kadi, A. S. (2004). FLASH, a component of the FAS-CAPSASE8 apoptotic pathway, is directly regulated by Hoxb4 in the notochord. *Dev. Biol.* **265**, 105-112.

Morsi El-Kadi, A. S., in der Reiden, P., Durston, A. and Morgan, R. (2002). The small GTPase Rap1 is an immediate downstream target for Hoxb4 transcriptional regulation. *Mech. Dev.* **113**, 131-139.

Neuteboom, S. T. and Murre, C. (1997). Pbx raises the DNA binding specificity but not the selectivity of antennapedia Hox proteins. *Mol Cell Biol* **17**, 4696-706.

Nikolajczyk, B. S., Nelsen, B. and Sen, R. (1996). Precise alignment of sites required for mu enhancer activation in B cells. *Mol. Cell. Biol.* **16**, 4544-4554.

Notredame, C., Higgins, D. G. and Heringa, J. (2000). T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol* **302**, 205-17.

Ostrowski, S., Dierick, H. A. and Bejsovec, A. (2002). Genetic control of cuticle formation during embryonic development of Drosophila melanogaster. *Genetics* **161**, 171-82.

Panganiban, G. and Rubenstein, J. L. (2002). Developmental functions of the Distal-less/Dlx homeobox genes. *Development* **129**, 4371-86.

Pearson, J. C., Lemons, D. and McGinnis, W. (2005). Modulating Hox gene functions during animal body patterning. *Nat Rev Genet* **6**, 893-904.

Pederson, J. A. (2000). Regulation by homeoproteins: a comparison of Deformed-responsive elements. *Genetics* **156**, 667-686.

Peifer, M. and Wieschaus, E. (1990). Mutations in the Drosophila gene extradenticle affect the way specific homeo domain proteins regulate segmental identity. *Genes Dev.* **4**, 1209-1223.

Pellerin, I., Schnabel, C., Catron, K. M. and Abate, C. (1994). Hox proteins have different affinities for a consensus DNA site that correlate with the positions of their genes on the hox cluster. *Mol Cell Biol* **14**, 4532-45.

Perkins, K. K., Dailey, G. M. and Tjian, R. (1988). Novel Jun- and Fos-related proteins in Drosophila are functionally homologous to enhancer factor AP-1. *Embo J* **7**, 4265-73.

Phillips, M. A., Jessen, B. A., Lu, Y., Qin, Q., Stevens, M. E. and Rice, R. H. (2004). A distal region of the human TGM1 promoter is required for expression in transgenic mice and cultured keratinocytes. *BMC Dermatol* **4**, 2.

Poliakov, A., Cotrina, M. and Wilkinson, D. G. (2004). Diverse roles of eph receptors and ephrins in the regulation of cell migration and tissue assembly. *Dev. Cell* **7**, 465-480.

Pollock, R. and Treisman, R. (1990). A sensitive method for the determination of protein-DNA binding specificities. *Nucleic Acids Res* **18**, 6197-204.

Popperl, H. (1995). Segmental expression of Hoxb-1 is controlled by a highly conserved autoregulatory loop dependent upon Exd/Pbx. *Cell* **81**, 1031-1042.

Ramet, M., Lanot, R., Zachary, D. and Manfruelli, P. (2002). JNK signaling pathway is required for efficient wound healing in Drosophila. *Dev Biol* **241**, 145-56.

Rauskolb, C., Smith, K., Peifer, M. and Wieschaus, E. (1995). extradenticle determines segmental identities throughout Drosophila development. *Development* **121**, 3663-3673.

Read, D., Nishigaki, T. and Manley, J. L. (1990). The Drosophila even-skipped promoter is transcribed in a stage-specific manner in vitro and contains multiple, overlapping factor-binding sites. *Mol Cell Biol* **10**, 4334-44.

Rebeiz, M. and Posakony, J. W. (2004). GenePalette: a universal software tool for genome sequence visualization and analysis. *Dev Biol* **271**, 431-8.

Rebeiz, M., Reeves, N. L. and Posakony, J. W. (2002). SCORE: a computational approach to the identification of cis-regulatory modules and target genes in whole-genome sequence data. Site clustering over random expectation. *Proc Natl Acad Sci U S A* **99**, 9888-93.

Rebeiz, M., Stone, T. and Posakony, J. W. (2005). An ancient transcriptional regulatory linkage. *Developmental Biology* **281**, 299-308.

Robertson, L. K., Bowling, D. B., Mahaffey, J. P., Imiolczyk, B. and Mahaffey, J. W. (2004). An interactive network of zinc-finger proteins contributes to regionalization of the Drosophila embryo and establishes the domains of HOM-C-protein function. *Development* **131**, 2781-2789.

Rodriguez-Trelles, F., Tarrio, R. and Ayala, F. J. (2003). Evolution of cis-regulatory regions versus codifying regions. *Int J Dev Biol* **47**, 665-73.

Rubin, G. M. and Spradling, A. C. (1982). Genetic transformation of Drosophila with transposable element vectors. *Science* **218**, 348-53.

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

Ryoo, H. D. and Mann, R. S. (1999). The control of trunk Hox specificity and activity by Extradenticle. *Genes Dev.* **13**, 1704-1716.

Safaei, R. (1997). A target of the HoxB5 gene from the mouse nervous system. *Brain Res. Dev. Brain Res.* **100**, 5-12.

Salser, S. and Kenyon, C. (1996). A C. elegans Hox gene switches on, off, on and off again to regulate proliferation, differentiation and morphogenesis. *Development* **122**, 1651-1661.

Salser, S. J. and Kenyon, C. (1992). Activation of a C. elegans Antennapedia homologue in migrating cells controls their direction of migration. *Nature* **355**, 255-258.

Schier, A. F. and Gehring, W. J. (1992). Direct homeodomain-DNA interaction in the autoregulation of the fushi tarazu gene. *Nature* **356**, 804-807.

Scholnick, S. B., Bray, S. J., Morgan, B. A., McCormick, C. A. and Hirsh, J. (1986). CNS and hypoderm regulatory elements of the Drosophila melanogaster dopa decarboxylase gene. *Science* **234**, 998-1002.

Schoppmeier, M. and Damen, W. G. (2001). Double-stranded RNA interference in the spider Cupiennius salei: the role of Distal-less is evolutionarily conserved in arthropod appendage formation. *Dev Genes Evol* **211**, 76-82.

Senger, K., Armstrong, G. W., Rowell, W. J., Kwan, J. M., Markstein, M. and Levine, M. (2004). Immunity regulatory DNAs share common organizational features in Drosophila. *Mol Cell* **13**, 19-32.

Serpente, P. (2005). Direct crossregulation between retinoic acid receptor [beta] and Hox genes during hindbrain segmentation. *Development* **132**, 503-513.

Sharrocks, A. D., Brown, A. L., Ling, Y. and Yates, P. R. (1997). The ETS-domain transcription factor family. *Int J Biochem Cell Biol* **29**, 1371-87.

Shi, X., Bai, S., Li, L. and Cao, X. (2001). Hoxa-9 represses transforming growth factor-[beta]-induced osteopontin gene transcription. *J. Biol. Chem.* **276**, 850-855.

Siddharthan, R., Siggia, E. D. and van Nimwegen, E. (2005). PhyloGibbs: a Gibbs sampling motif finder that incorporates phylogeny. *PLoS Comput Biol* **1**, e67.

Sinha, S., Blanchette, M. and Tompa, M. (2004). PhyME: a probabilistic algorithm for finding motifs in sets of orthologous sequences. *BMC Bioinformatics* **5**, 170.

Sonnhammer, E. L. and Durbin, R. (1995). A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. *Gene* **167**, 10.

Spradling, A. C. and Rubin, G. M. (1982). Transposition of cloned P elements into Drosophila germ line chromosomes. *Science* **218**, 341-7.

Spradling, A. C., Stern, D., Beaton, A., Rhem, E. J., Laverty, T., Mozden, N., Misra, S. and Rubin, G. M. (1999). The Berkeley Drosophila Genome Project gene disruption project: Single P-element insertions mutating 25% of vital Drosophila genes. *Genetics* **153**, 135-77.

Stadler, H. S., Higgins, K. M. and Capecchi, M. R. (2001). Loss of Eph-receptor expression correlates with loss of cell adhesion and chondrogenic capacity in Hoxa13 mutant limbs. *Development* **128**, 4177-4188.

Stramer, B., Wood, W., Galko, M. J., Redd, M. J., Jacinto, A., Parkhurst, S. M. and Martin, P. (2005). Live imaging of wound inflammation in Drosophila embryos reveals key roles for small GTPases during in vivo cell migration. *J Cell Biol* **168**, 567-73.

Streit, A. (2002). Conserved regulation of the Caenorhabditis elegans labial/Hox1 gene ceh-13. *Dev. Biol.* **242**, 96-108.

Strutt, D. I. and White, R. A. (1994). Characterization of T48, a target of homeotic gene regulation in Drosophila embryogenesis. *Mech. Dev.* **46**, 27-39.

Su, Y. C., Treisman, J. E. and Skolnik, E. Y. (1998). The Drosophila Ste20-related kinase misshapen is required for embryonic dorsal closure and acts through a JNK MAPK module on an evolutionarily conserved signaling pathway. *Genes Dev* **12**, 2371-80.

Sucena, E., Delon, I., Jones, I., Payre, F. and Stern, D. L. (2003). Regulatory evolution of shavenbaby/ovo underlies multiple cases of morphological parallelism. *Nature* **424**, 935-8.

Sucena, E. and Stern, D. L. (2000). Divergence of larval morphology between Drosophila sechellia and its sibling species caused by cis-regulatory evolution of ovo/shaven-baby. *Proc Natl Acad Sci U S A* **97**, 4530-4.

Sun, B., Hursh, D. A., Jackson, D. and Beachy, P. A. (1995). Ultrabithorax protein is necessary but not sufficient for full activation of decapentaplegic expression in the visceral mesoderm. *EMBO J.* **14**, 520-535.

Sutter, N. B., Bustamante, C. D., Chase, K., Gray, M. M., Zhao, K., Zhu, L., Padhukasahasram, B., Karlins, E., Davis, S., Jones, P. G. et al. (2007). A Single IGF1 Allele Is a Major Determinant of Small Size in Dogs. *Science* **316**, 112-115.

Tagle, D. A., Koop, B. F., Goodman, M., Slightom, J. L., Hess, D. L. and Jones, R. T. (1988). Embryonic epsilon and gamma globin genes of a prosimian primate (Galago crassicaudatus). Nucleotide and amino acid sequences, developmental regulation and phylogenetic footprints. *J Mol Biol* **203**, 439-55.

Tamura, K., Subramanian, S. and Kumar, S. (2004). Temporal Patterns of Fruit Fly (Drosophila) Evolution Revealed by Mutation Clocks. *Molecular Biology and Evolution* **21**, 36-44.

Theokli, C., Morsi El-Kadi, A. S. and Morgan, R. (2003). TALE class homeodomain gene Irx5 is an immediate downstream target for Hoxb4 transcriptional regulation. *Dev. Dyn.* **227**, 48-55.

Thorsteinsdottir, U. (1997). Overexpression of HOXA10 in murine hematopoietic cells perturbs both myeloid and lymphoid differentiation and leads to acute myeloid leukemia. *Mol. Cell. Biol.* **17**, 495-505.

Ting, S. B., Caddy, J., Hislop, N., Wilanowski, T., Auden, A., Zhao, L. L., Ellis, S., Kaur, P., Uchida, Y., Holleran, W. M. et al. (2005). A homolog of Drosophila grainy head is essential for epidermal integrity in mice. *Science* **308**, 411-3.

True, J. R. and Haag, E. S. (2001). Developmental system drift and flexibility in evolutionary trajectories. *Evolution and Development* **3**, 109-119.

Uv, A. E., Harrison, E. J. and Bray, S. J. (1997). Tissue-specific splicing and functions of the Drosophila transcription factor Grainyhead. *Mol Cell Biol* **17**, 6727-35.

Uv, A. E., Thompson, C. R. and Bray, S. J. (1994). The Drosophila tissue-specific factor Grainyhead contains novel DNA-binding and dimerization domains which are conserved in the human protein CP2. *Mol Cell Biol* **14**, 4020-31.

Vachon, G. (1992). Homeotic genes of the bithorax complex repress limb development in the abdomen of the Drosophila embryo through the target gene Distal-less. *Cell* **71**, 437-450.

Vachon, G., Cohen, B., Pfeifle, C., McGuffin, M. E., Botas, J. and Cohen, S. M. (1992). Homeotic genes of the Bithorax complex repress limb development in the abdomen of the Drosophila embryo through the target gene Distal-less. *Cell* **71**, 437-50.

Van Auken, K. (2002). Roles of the Homothorax/Meis/Prep homolog UNC-62 and the Exd/Pbx homologs CEH-20 and CEH-40 in C. elegans embryogenesis. *Development* **129**, 5255-5268.

van Dijk, M. A. and Murre, C. (1994). extradenticle Raises the DNA binding specificity of homeotic selector gene products. *Cell* **78**, 617-624.

van Dijk, M. A. V., Voorhoeve, P. M. and Murre, C. (1993). Pbx1 is Converted into a Transcriptional Activator Upon Acquiring the N-Terminal Region of E2A in Pre-B-Cell Acute Lymphoblastoid Leukemia. *PNAS* **90**, 6061-6065.

Venkatesan, K., McManus, H. R., Mello, C. C., Smith, T. F. and Hansen, U. (2003). Functional conservation between members of an ancient duplicated transcription factor family, LSF/Grainyhead. *Nucleic Acids Res* **31**, 4304-16.

Vincent, J. and Wegst, U. (2004). Design and mechanical properties of insect cuticle. *Arthropod Structure & Development* **33**, 187-199.

Wang, B. B. (1993). A homeotic gene cluster patterns the anteroposterior body axis of C. elegans. *Cell* **74**, 29-42.

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

White, R. A., Aspland, S. E., Brookman, J. J., Clayton, L. and Sproat, G. (2000). The design and analysis of a homeotic response element. *Mech Dev* **91**, 217-26.

Williams-Masson, E. M., Malik, A. N. and Hardin, J. (1997). An actin-mediated two-step mechanism is required for ventral enclosure of the C. elegans hypodermis. *Development* **124**, 2889-901.

Williams, T. M. (2005). Candidate downstream regulated genes of HOX group 13 transcription factors with and without monomeric DNA binding capability. *Dev. Biol.* **279**, 462-480.

Wittkopp, P. J. (2006). Evolution of cis-regulatory sequence and function in Diptera. *Heredity* **97**, 139-47.

Wood, W., Jacinto, A., Grose, R., Woolner, S., Gale, J., Wilson, C. and Martin, P. (2002). Wound healing recapitulates morphogenesis in Drosophila embryos. *Nat Cell Biol* **4**, 907-12.

Wratten, N. S., McGregor, A. P., Shaw, P. J. and Dover, G. A. (2006). Evolutionary and functional analysis of the tailless enhancer in Musca domestica and Drosophila melanogaster. *Evol Dev* **8**, 6-15.

Wray, G. A., Hahn, M. W., Abouheif, E., Balhoff, J. P., Pizer, M., Rockman, M. V. and Romano, L. A. (2003). The evolution of transcriptional regulation in eukaryotes. *Mol Biol Evol* **20**, 1377-419.

Xiong, B. and Jacobs-Lorena, M. (1995). Gut-specific transcriptional regulatory elements of the carboxypeptidase gene are conserved between black flies and Drosophila. *Proc Natl Acad Sci U S A* **92**, 9313-7.

Yates, S. and Rayner, T. E. (2002). Transcription factor activation in response to cutaneous injury: role of AP-1 in reepithelialization. *Wound Repair Regen* **10**, 5-15.

Yokouchi, Y. (1995). Misexpression of Hoxa-13 induces cartilage homeotic transformation and changes cell adhesiveness in chick limb buds. *Genes Dev.* **9**, 2509-2522.

Young, P. E., Richman, A. M., Ketchum, A. S. and Kiehart, D. P. (1993). Morphogenesis in Drosophila requires nonmuscle myosin heavy chain function. *Genes Dev* **7**, 29-41.

Yu, Z., Lin, K. K., Bhandari, A., Spencer, J. A., Xu, X., Wang, N., Lu, Z., Gill, G. N., Roop, D. R., Wertz, P. et al. (2006). The Grainyhead-like epithelial transactivator Get-1/Grhl3 regulates epidermal terminal differentiation and interacts functionally with LMO4. *Dev Biol* **299**, 122-36.

Zaffran, S., Kuchler, A., Lee, H. H. and Frasch, M. (2001). biniou (FoxF), a central component in a regulatory network controlling visceral mesoderm development and midgut morphogenesis in Drosophila. *Genes Dev.* **15**, 2900-2915.

Zakany, J. and Duboule, D. (1999). Hox genes in digit development and evolution. *Cell Tissue Res.* **296**, 19-25.

Zeng, C., Pinsonneault, J., Gellon, G., McGinnis, N. and McGinnis, W. (1994). Deformed protein binding sites and cofactor binding sites are required for the function of a small segment-specific regulatory element in Drosophila embryos. *EMBO J.* **13**, 2362-2377.

Zhang, K., Chaillet, J. R., Perkins, L. A., Halazonetis, T. D. and Perrimon, N. (1990). Drosophila homolog of the mammalian jun oncogene is expressed during embryonic development and activates transcription in mammalian cells. *Proc Natl Acad Sci U S A* **87**, 6281-5.

Zhou, B., Bagri, A. and Beckendorf, S. K. (2001). Salivary gland determination in Drosophila: a salivary-specific, fork head enhancer integrates spatial pattern and allows fork head autoregulation. *Dev. Biol.* **237**, 54-67.