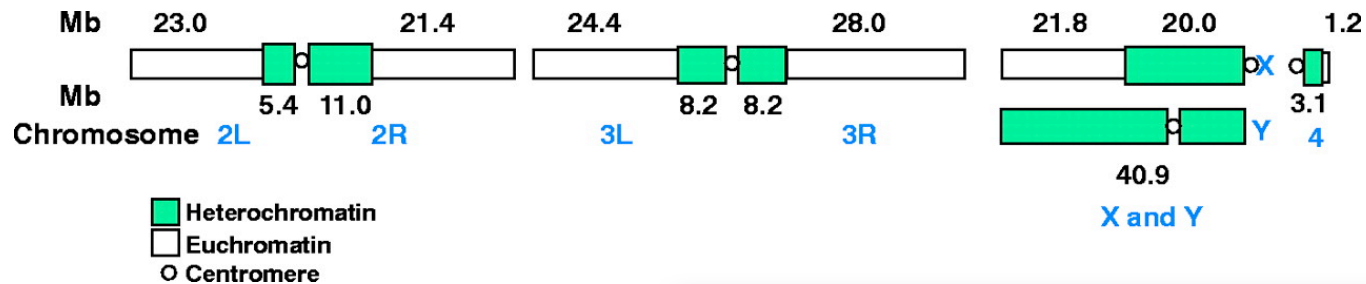
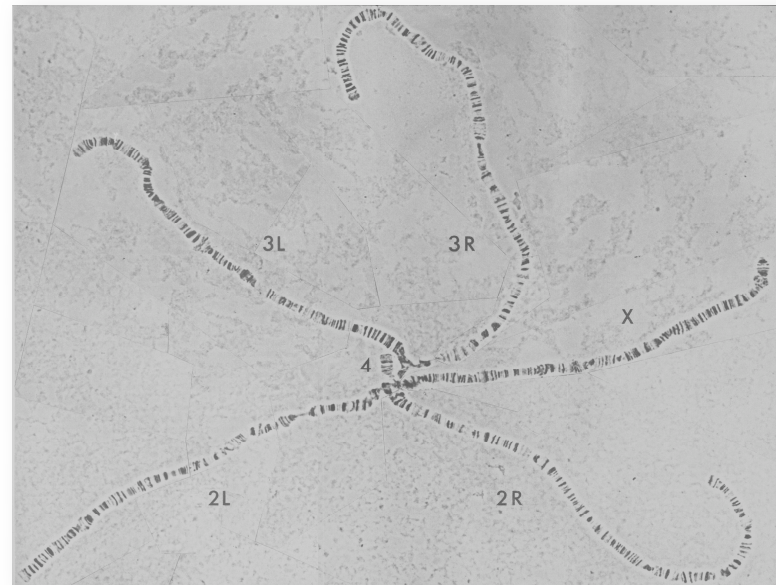


# The *D. melanogaster* genome



~120 Mb of euchromatin  
~15,000 genes  
75% noncoding



# The function of most noncoding DNA is unknown & unannotated

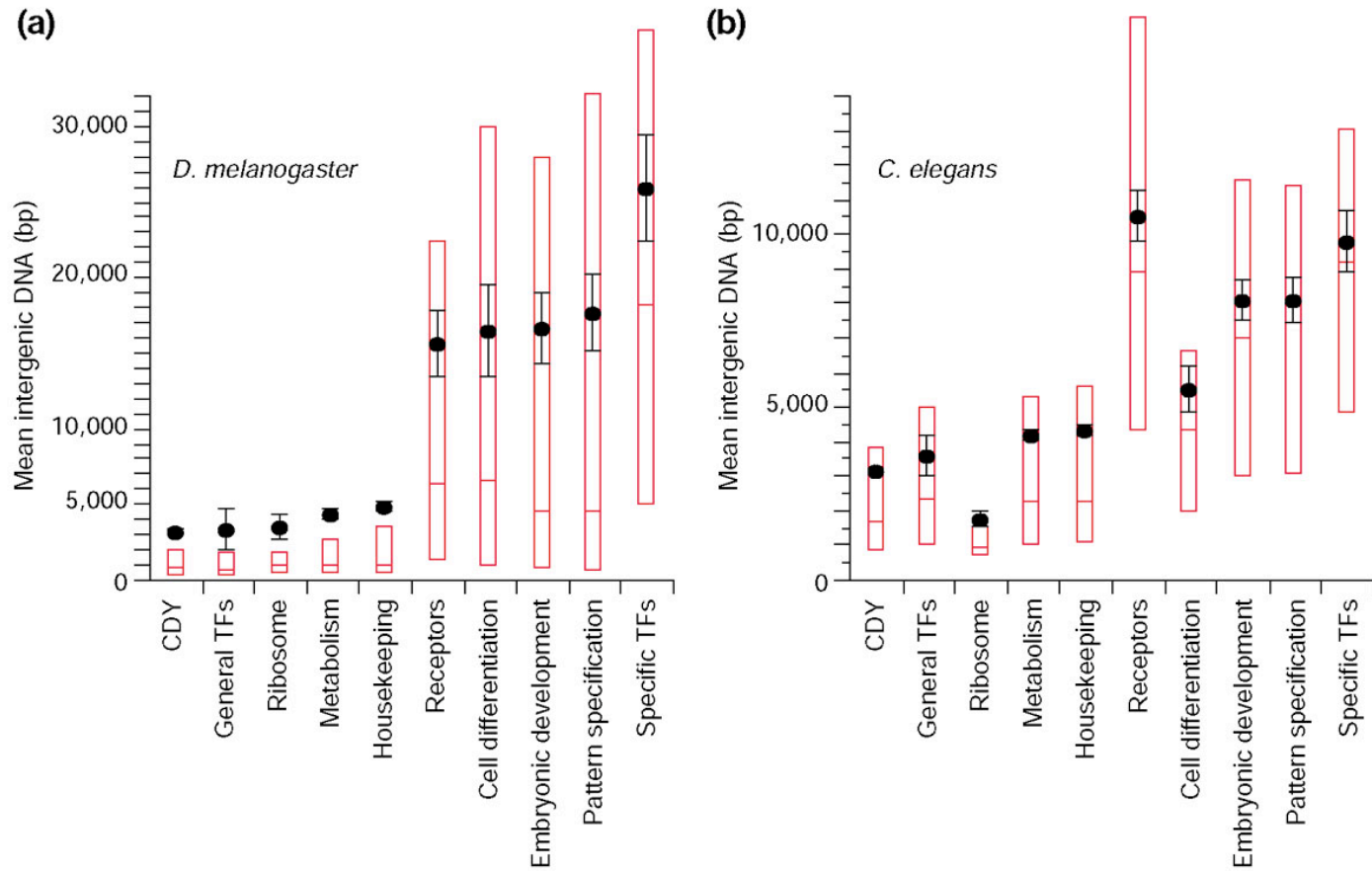


Bioinformatic & functional analysis of noncoding DNA ⇒

Genome organization

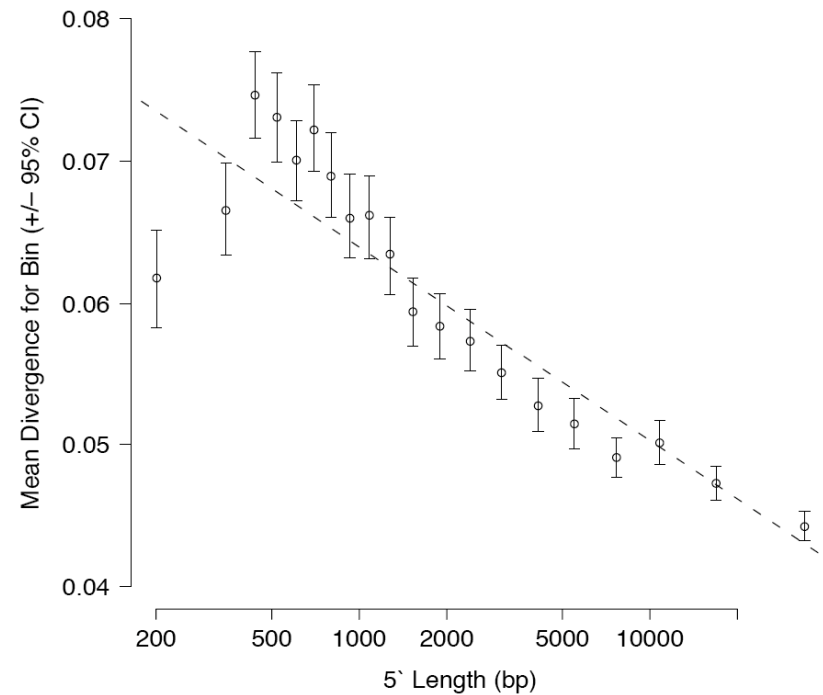
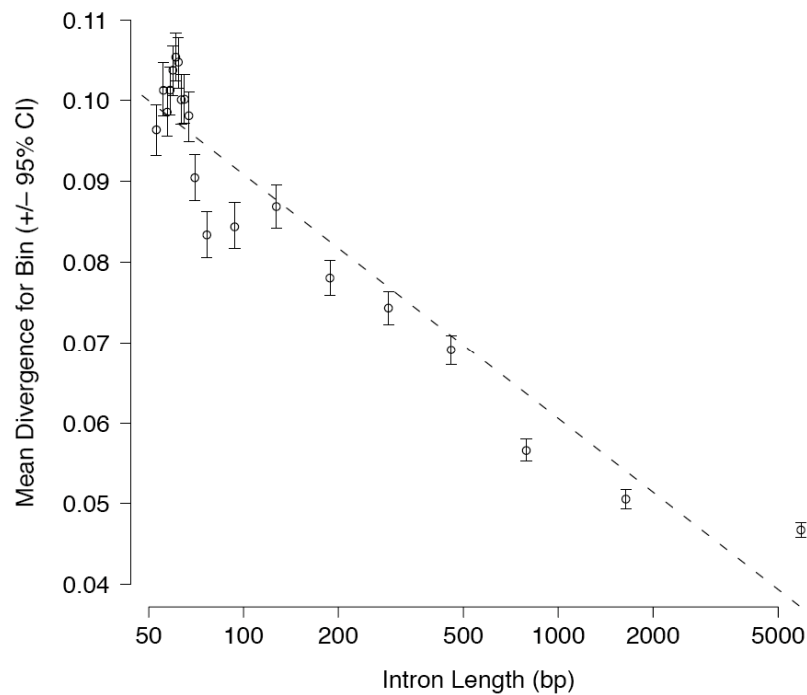
Transcriptional regulation

# Genes with complex expression have long intergenic regions in compact genomes



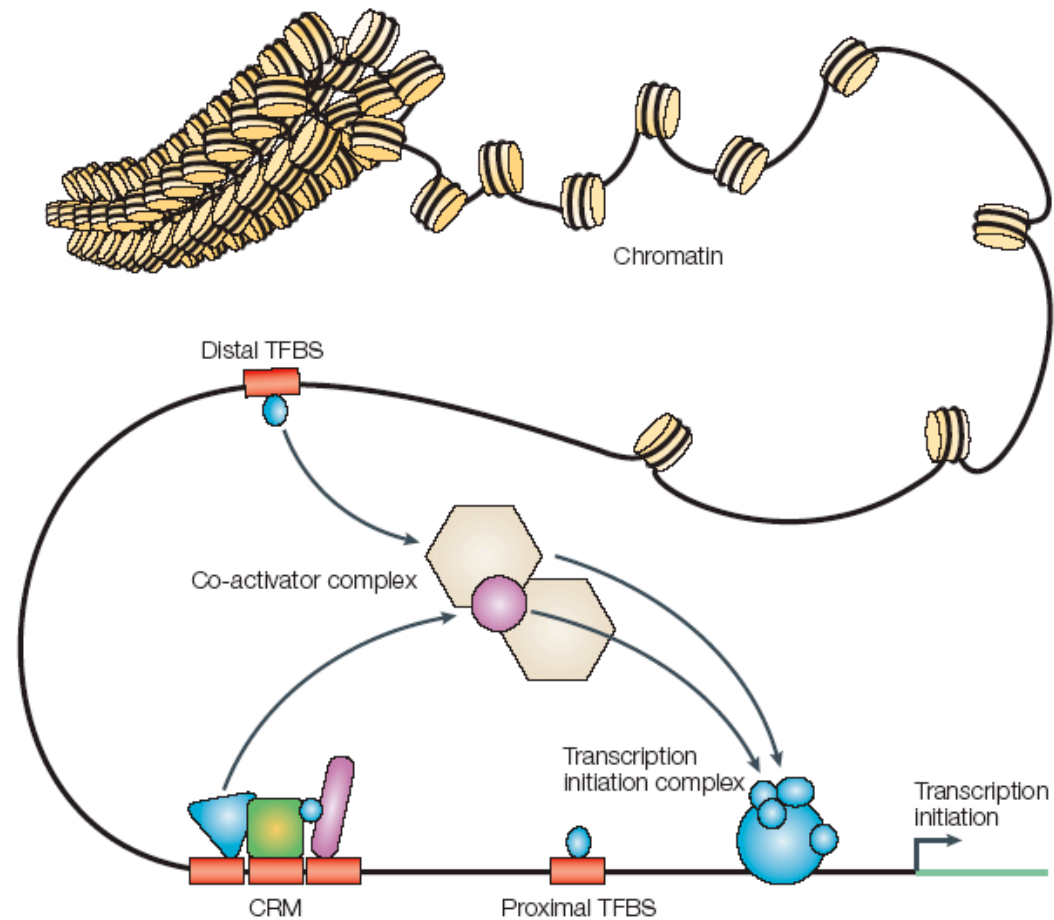
Nelson, Hersh & Carroll (2004) *Genome Biology* 5:R25

# Long introns & intergenic regions have slower rates of sequence evolution in *Drosophila*

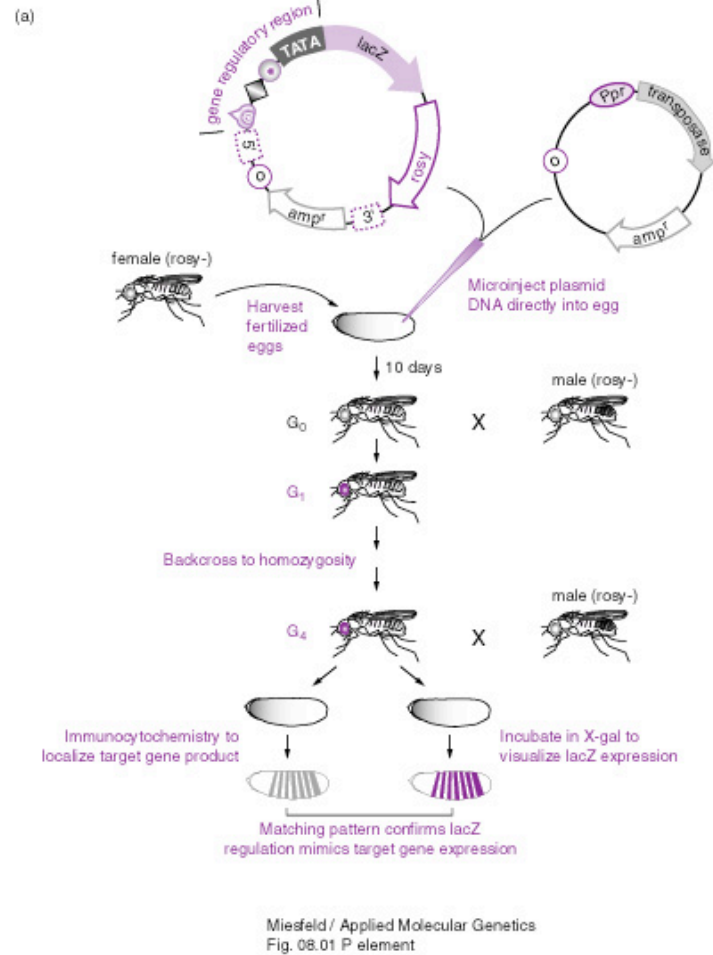
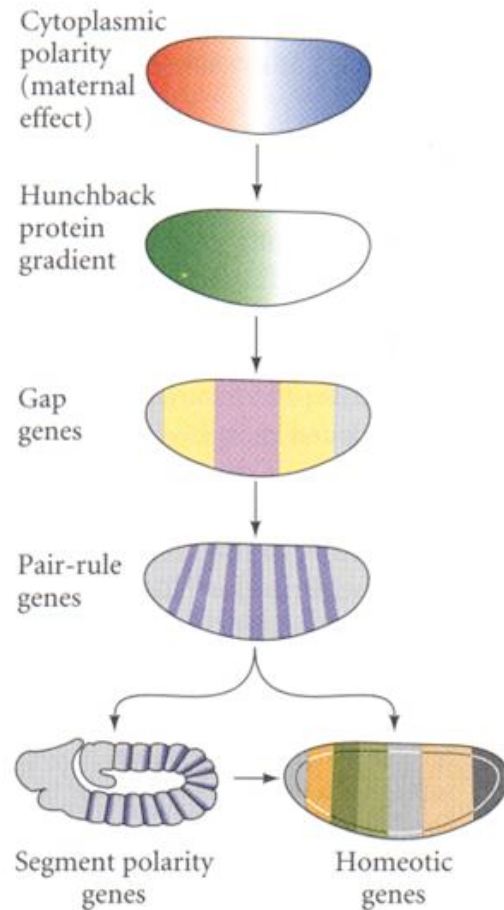


Halligan & Keightley (2006) *Genome Research* 16:875-884

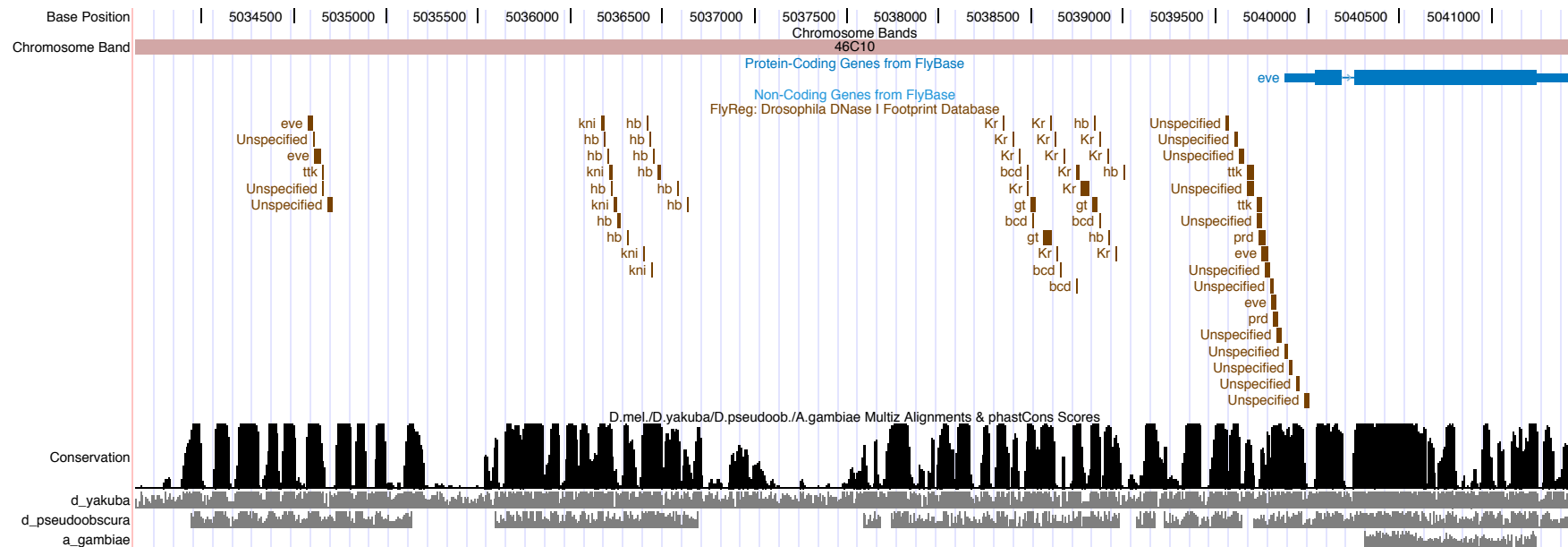
# Prediction: the vast majority of *Drosophila* noncoding DNA is *cis*-regulatory sequence



# *Drosophila* revealed the fundamental importance of gene regulation in developmental genetics



# Systematic annotation of *cis*-regulatory data in *Drosophila*: FlyReg & REDfly databases

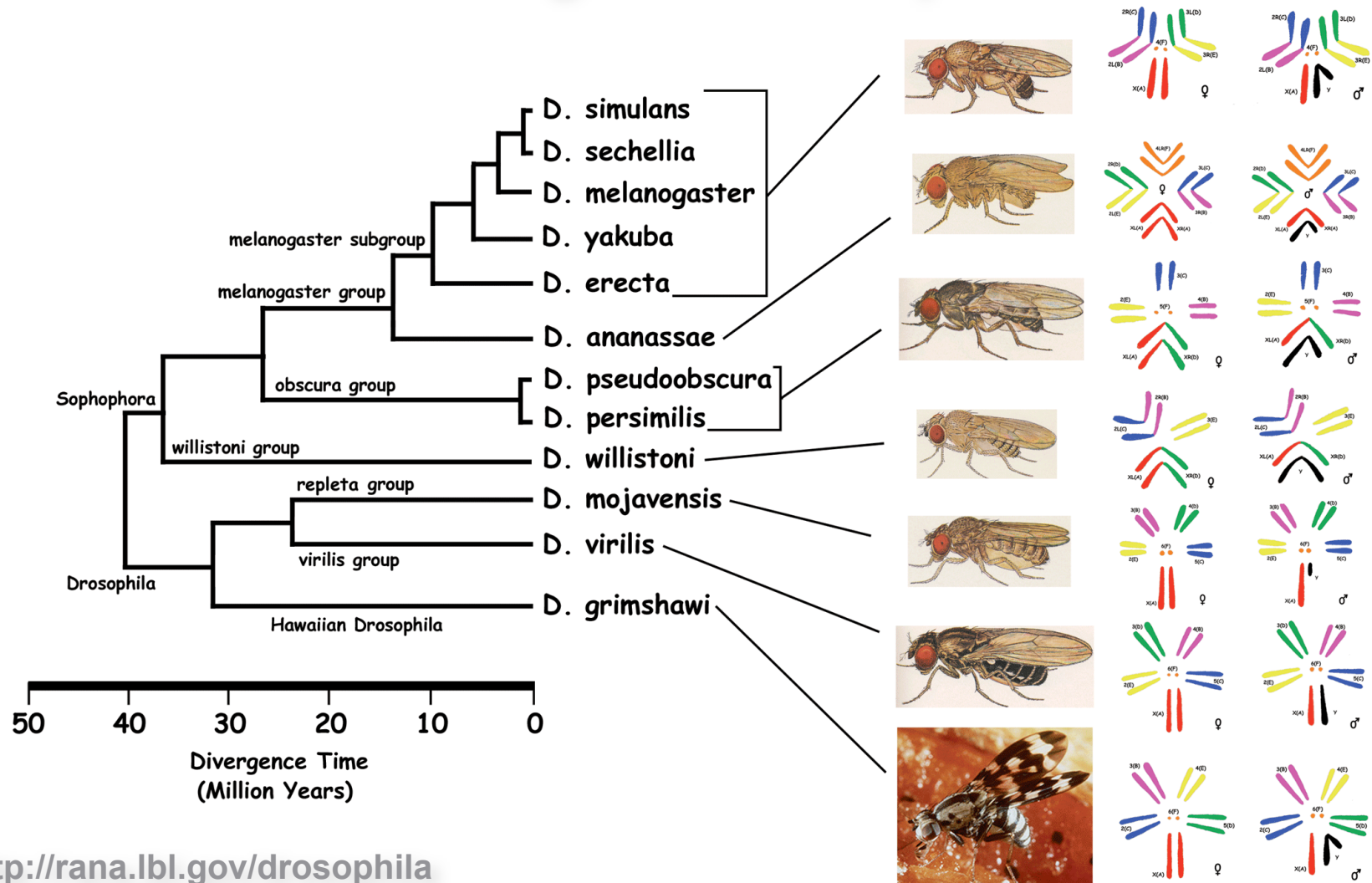


Bergman *et al.* (2005) *Bioinformatics* 21:1747-1749 (n=1365)

Gallo *et al.* (2006) *Bioinformatics* 22:381-383 (n=628)

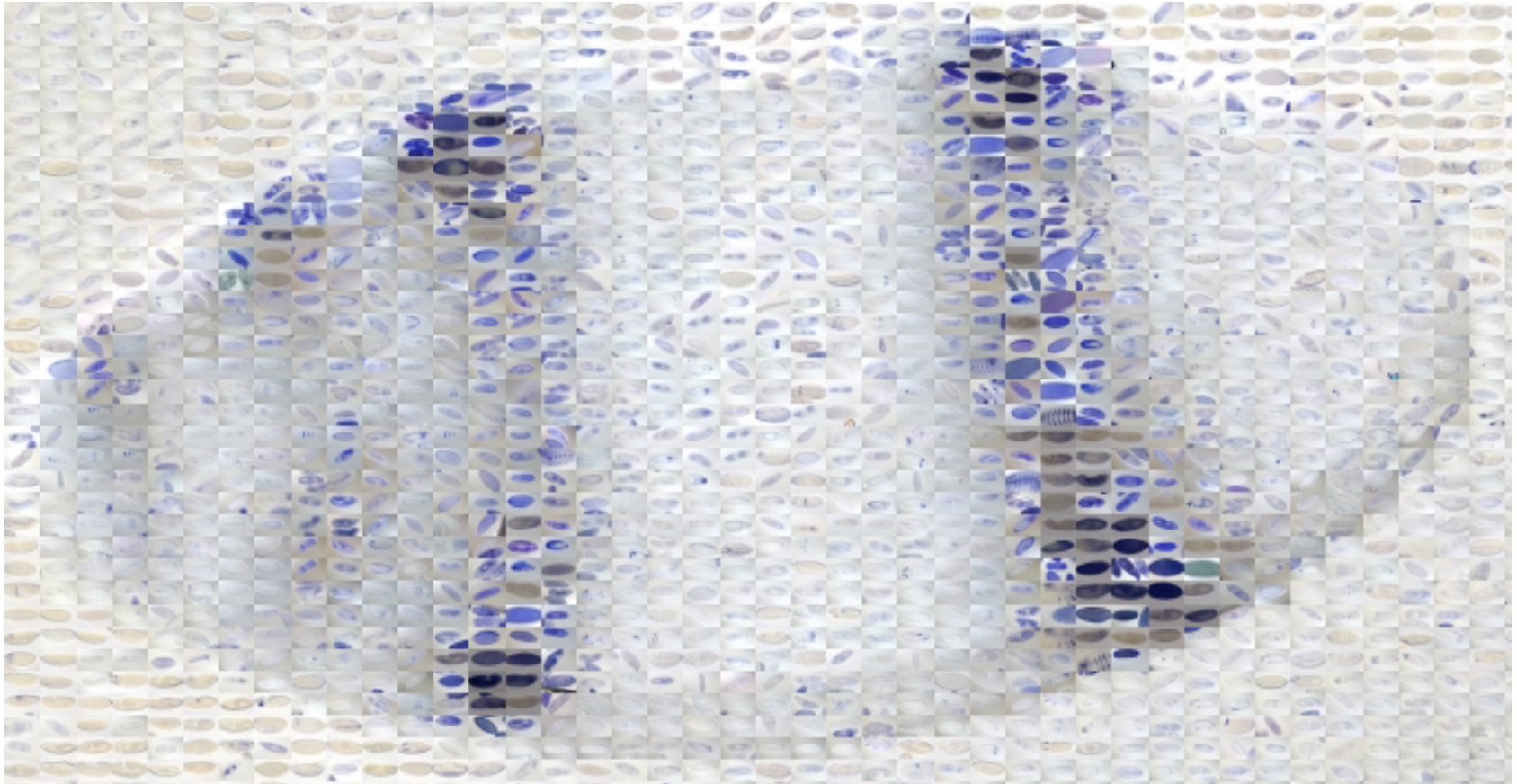


# A wealth of comparative genomic data exists for the genus *Drosophila*



<http://rana.lbl.gov/drosophila>  
<http://species.flybase.net>

# Thousands of candidate expression patterns: BDGP embryonic *in situ* database



<http://www.fruitfly.org/cgi-bin/ex/insitu.pl>

image from Pavel Tomancak (MPI-Dresden)

# Major promoter elements in *Drosophila*

## 1) TATA - box

- a. Hogness-Goldberg box
- b. -25/-30 upstream of TSS, recruits TBP
- c. not present in all *Drosophila* genes

## 2) INR (initiator element)

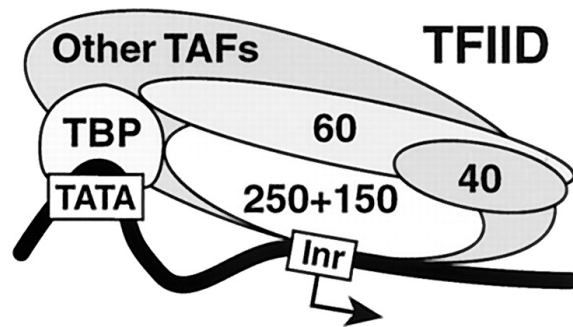
- a. found at TSS cap-site (-10/+10)
- b. found in 25% arthropod promoters
- c. not restricted to TATA-less

## 3) DPE (downstream promoter element)

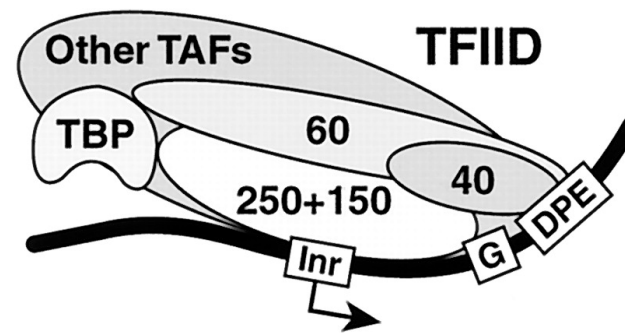
- a. +20/+30 downstream of TSS
- b. found first in *Drosophila* retrotransposons
- c. not restricted to TATA-less

# Different promoter configurations in *Drosophila*

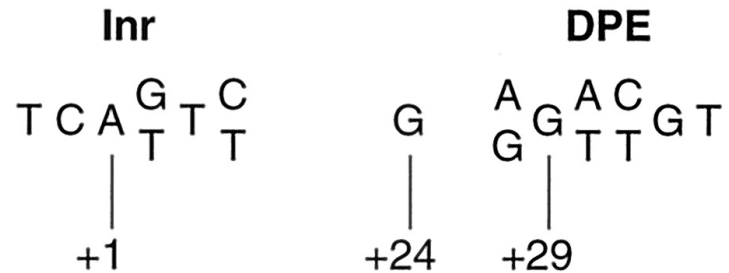
TATA-driven promoter



DPE-driven promoter



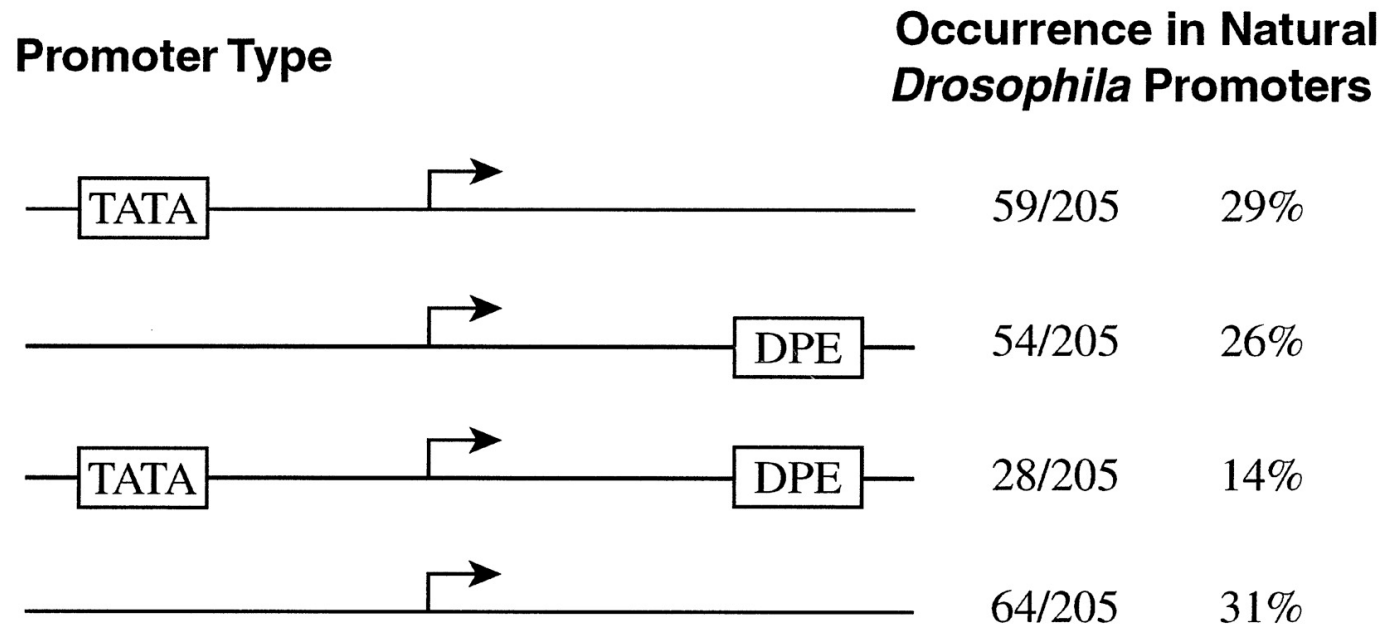
## Revised DPE Consensus



Kutach & Kadonaga (2000) *Mol. Cell Biol.* 20: 4754-4764

# Use of major promoter elements

**A**













Kutach & Kadonaga (2000) *Mol. Cell Biol.* 20: 4754-4764  
<<http://www-biology.ucsd.edu/labs/Kadonaga/DCPD.htm>>

# *Drosophila* core promoter elements can be reliably learned from genomic DNA

Table 2

The ten most significant motifs in the core promoter sequences from -60 to +40, as identified by the MEME algorithm

Motif	Pictogram	Bits	Consensus	Number	E value
1		15.2	YGGTCACACTR	311	5.1e-415
2 DRE		13.3	WATCGATW	277	1.7e-183
3 TATA		13.2	STATAWAAR	251	2.1e-138
4 INR		11.6	TCAGTYKNNNTYNR	369	3.4e-117
5		15.2	AWCAGCTGWT	125	2.9e-93
6		15.1	KTYRGATWTTT	107	1.9e-62
7		12.7	KNNCAKCNCTRNY	197	1.9e-63
8		14.7	MKSYGGCARGCSYSS	82	5.1e-29
9 DPE		15.4	CRWMGCGWKCGGTTS	56	1.9e-12
10		15.3	CSARCSSAACGS	40	8.3e-9

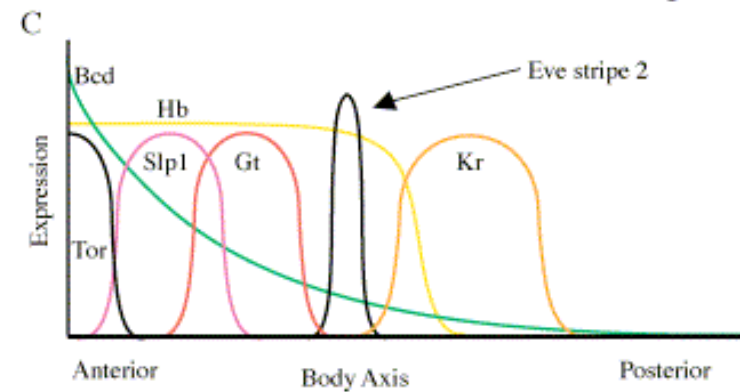
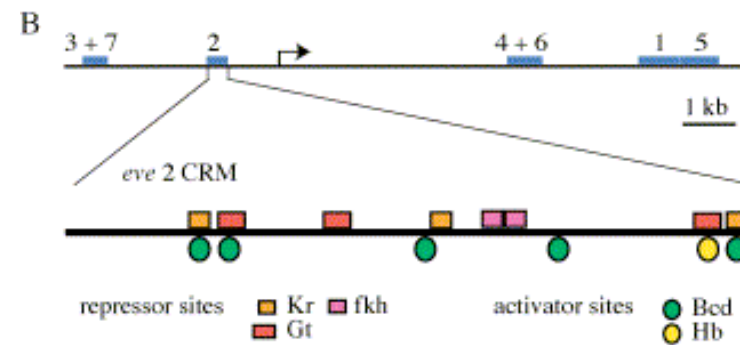
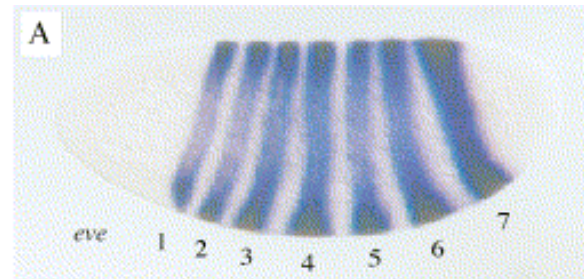
Ohler et al (2002) *Genome Biol.* 20: 4754-4764

# *Drosophila* core promoter elements can be reliably learned from genomic DNA

	Inferred Motif	Divergence	Ohler et al. Motif	
TIFDMEM0000116		0.13		Motif 1
TIFDMEM0000079		0.25		Motif 5 (E-box)
TIFDMEM0000005		0.26		Motif 2 (DRE)
TIFDMEM0000091		0.32		Motif 6
TIFDMEM0000083		0.47		Motif 3 (TATA)
TIFDMEM0000042		0.83		Motif 7
TIFDMEM0000036		1.31		Motif 4 (INR)
TIFDMEM0000057		2.58		Motif 8

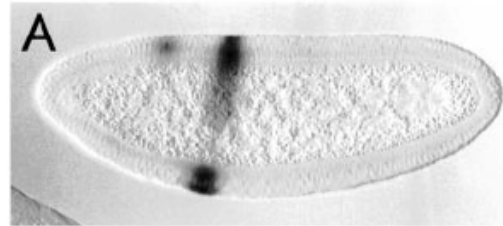
Down, Bergman, Su & Hubbard (2007) *PLoS Comp Biol* 3:e7.

# eve stripe 2: a model eukaryotic enhancer





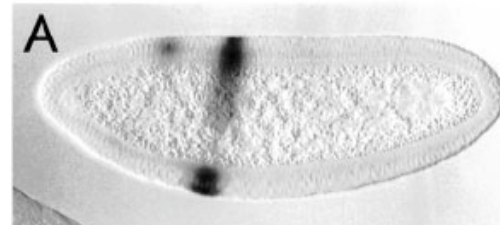
# eve stripe 2: a model eukaryotic enhancer



- + *bicoid*
- + *hunchback*
- - *Kruppel*
- - *giant*



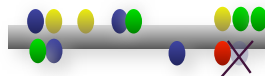
# eve stripe 2: a model eukaryotic enhancer



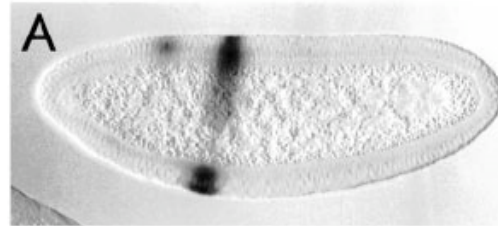
- + *bicoid*
- + *hunchback*
- - *Kruppel*
- - *giant*



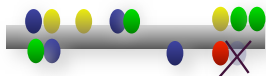
mutagenesis



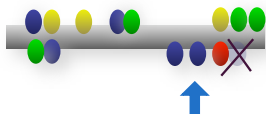
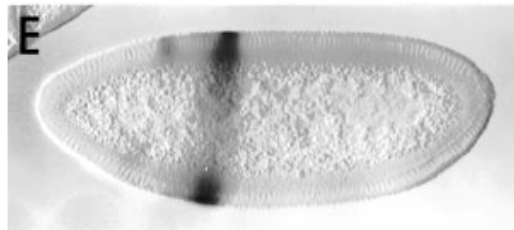
# eve stripe 2: a model eukaryotic enhancer



- + *bicoid*
- + *hunchback*
- - *Kruppel*
- - *giant*



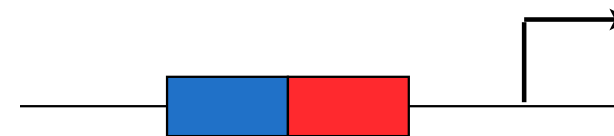
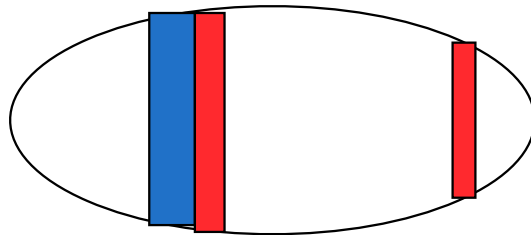
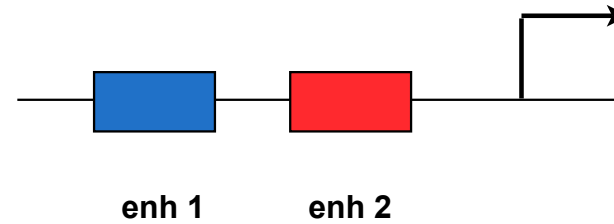
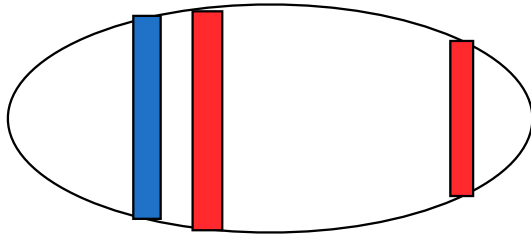
mutagenesis



*cis* complementation

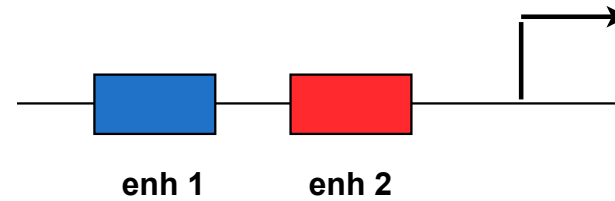
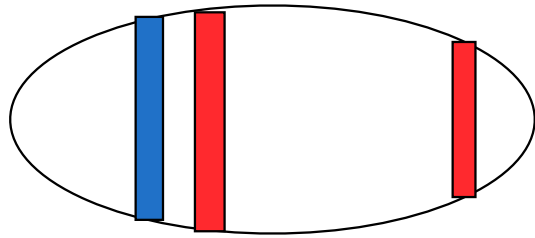
Arnosti et al. (1996) *Development* 122:205-214

# Spacing ensures enhancer autonomy

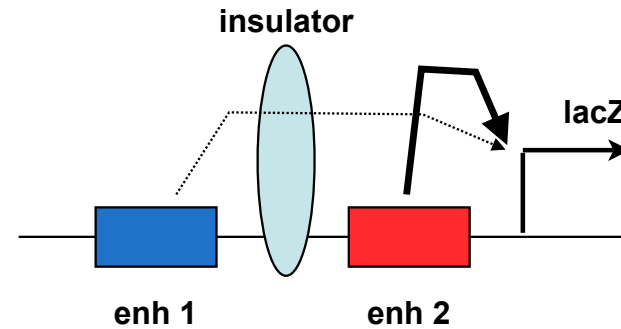
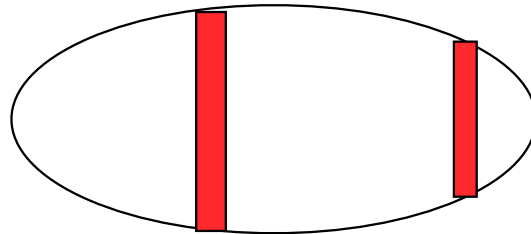


see Small et al (1996) *Development* 119: 767-772

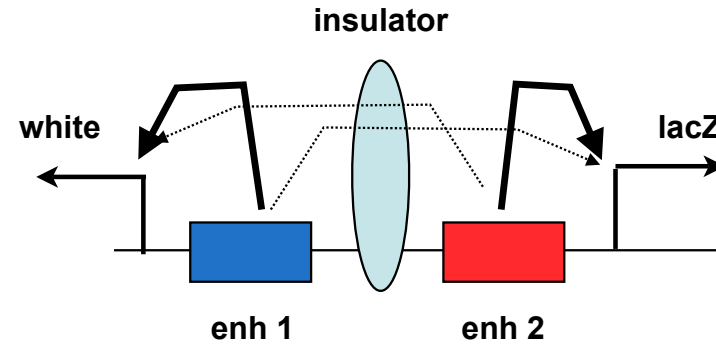
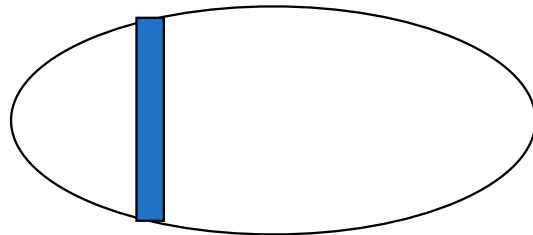
# Insulators can block *distal* enhancer-promoter interactions



*in situ*  
for **lacZ**  
mRNA



*in situ*  
for **white**  
mRNA

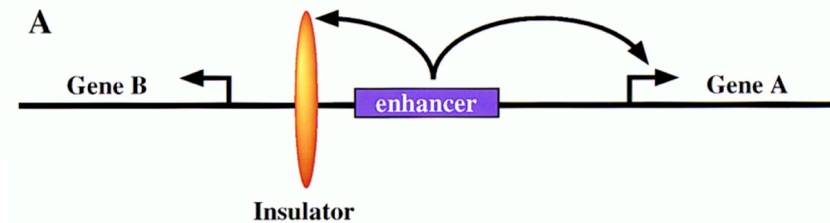
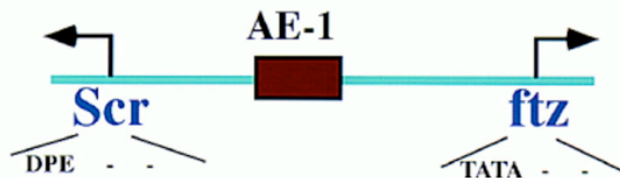
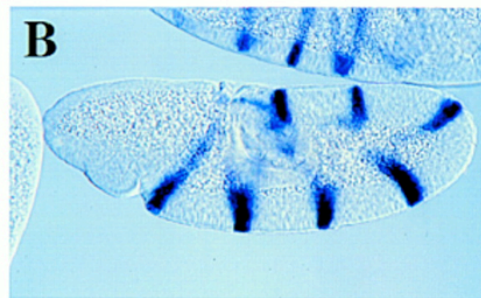
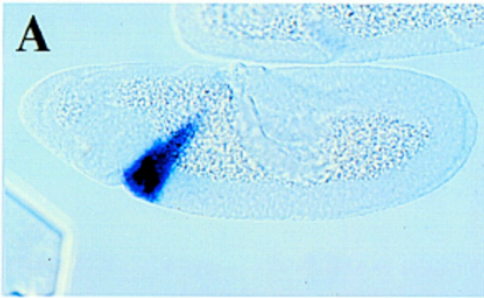


see Cai & Levine (1995) *Nature* 376:533-536

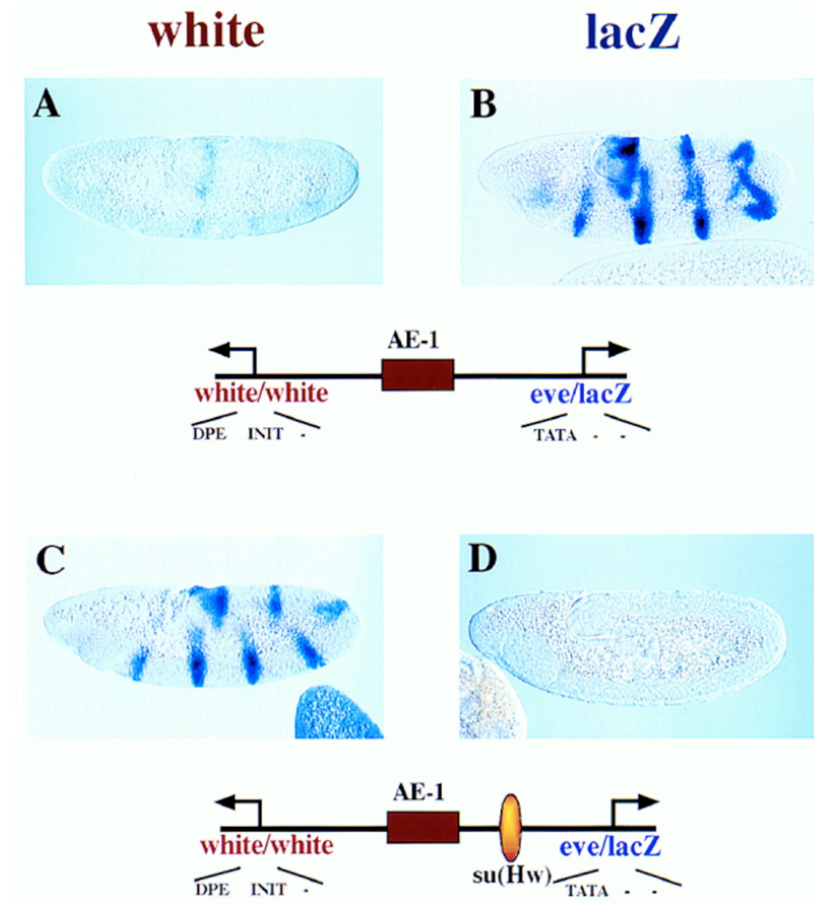
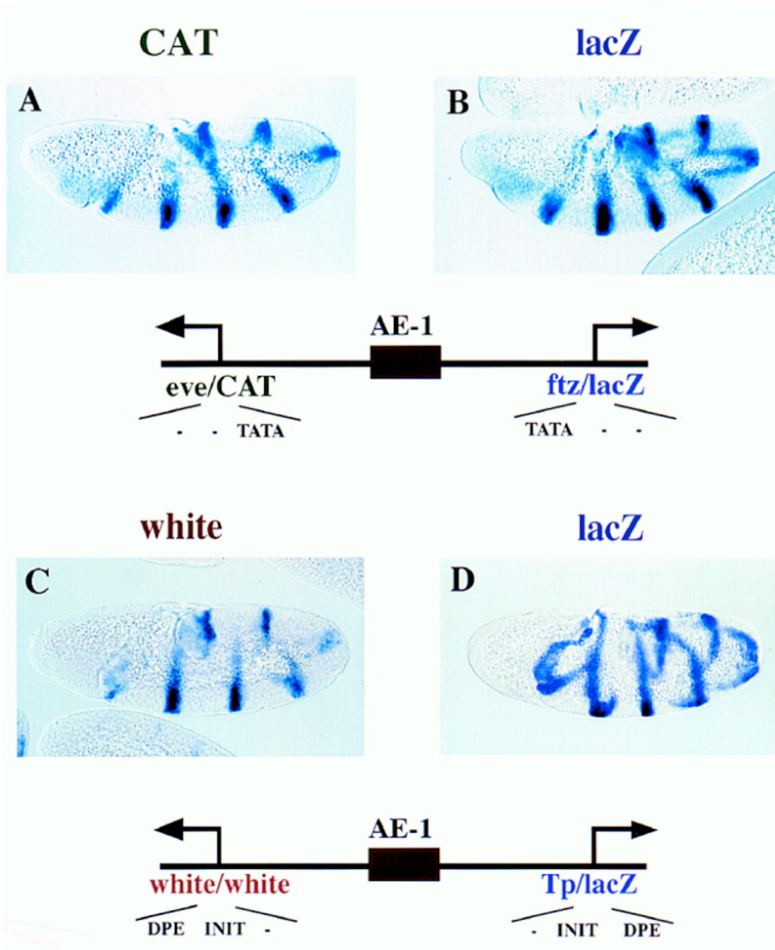
# What mechanisms ensure enhancer-promoter specificity *in vivo*?

Scr

ftz

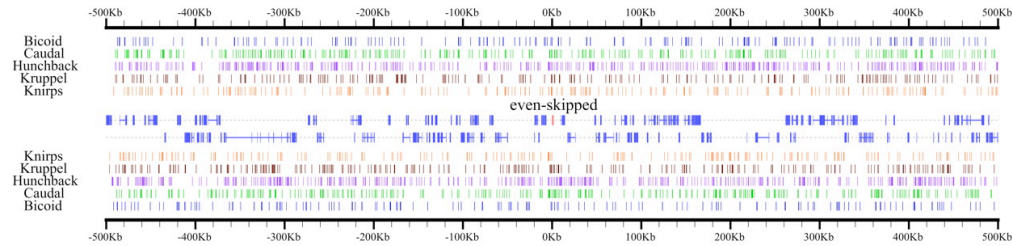


# Core promoter sequences can regulate enhancer-promoter interactions

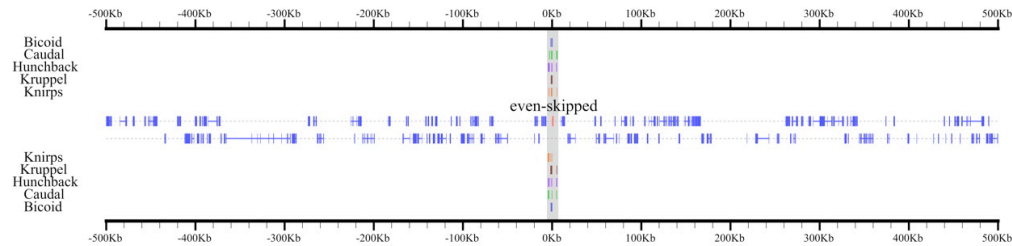


# CRMs can be predicted using binding site clustering

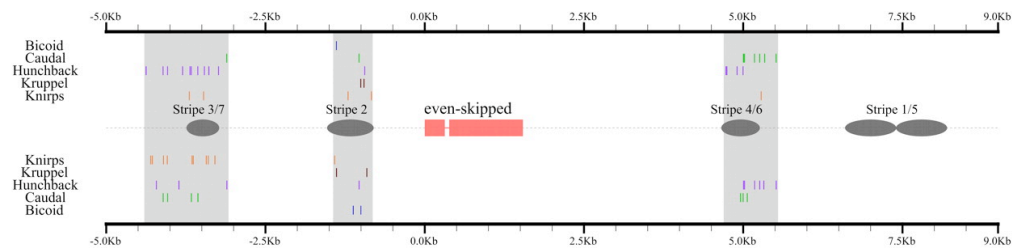
(A) High stringency matches



(B) High stringency matches and clustering filter



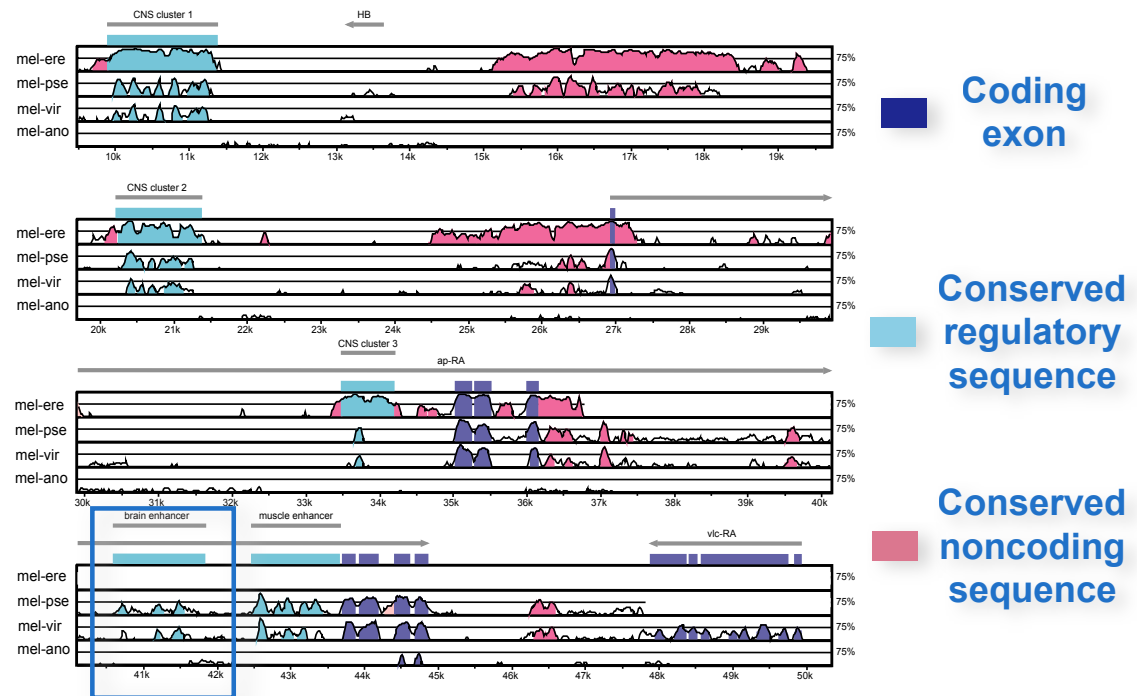
(C) Expanded view of *even-skipped* region



Berman et al. (2002) *PNAS* 22:757-762

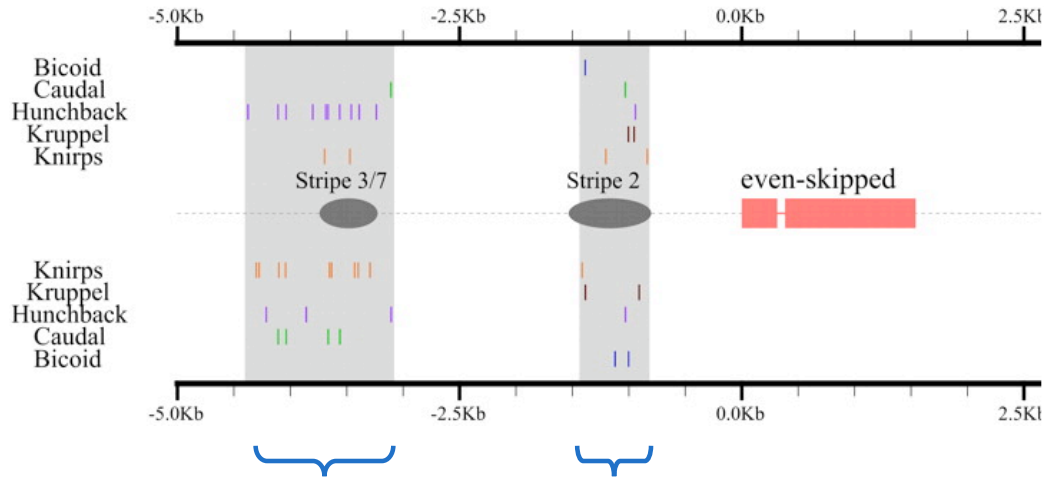


# CRMs can be predicted using clusters of conserved noncoding sequences

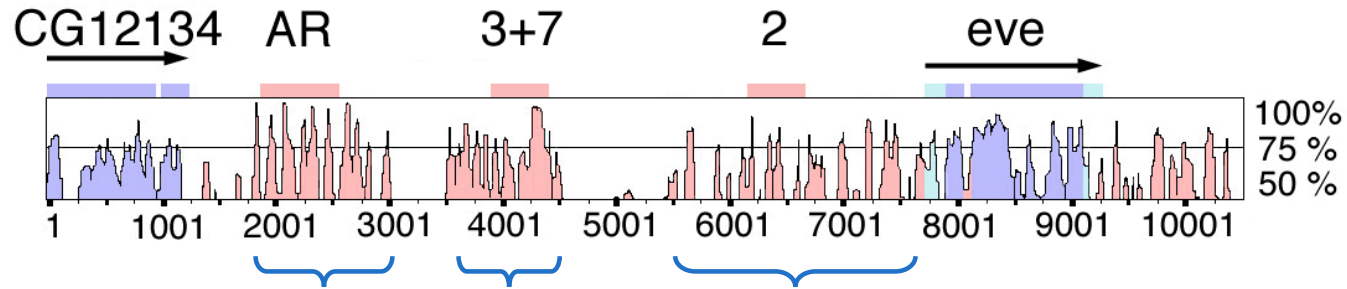


Bergman *et al.* (2002) *Genome Biology* 3:0086.

# Two views on CRM detection

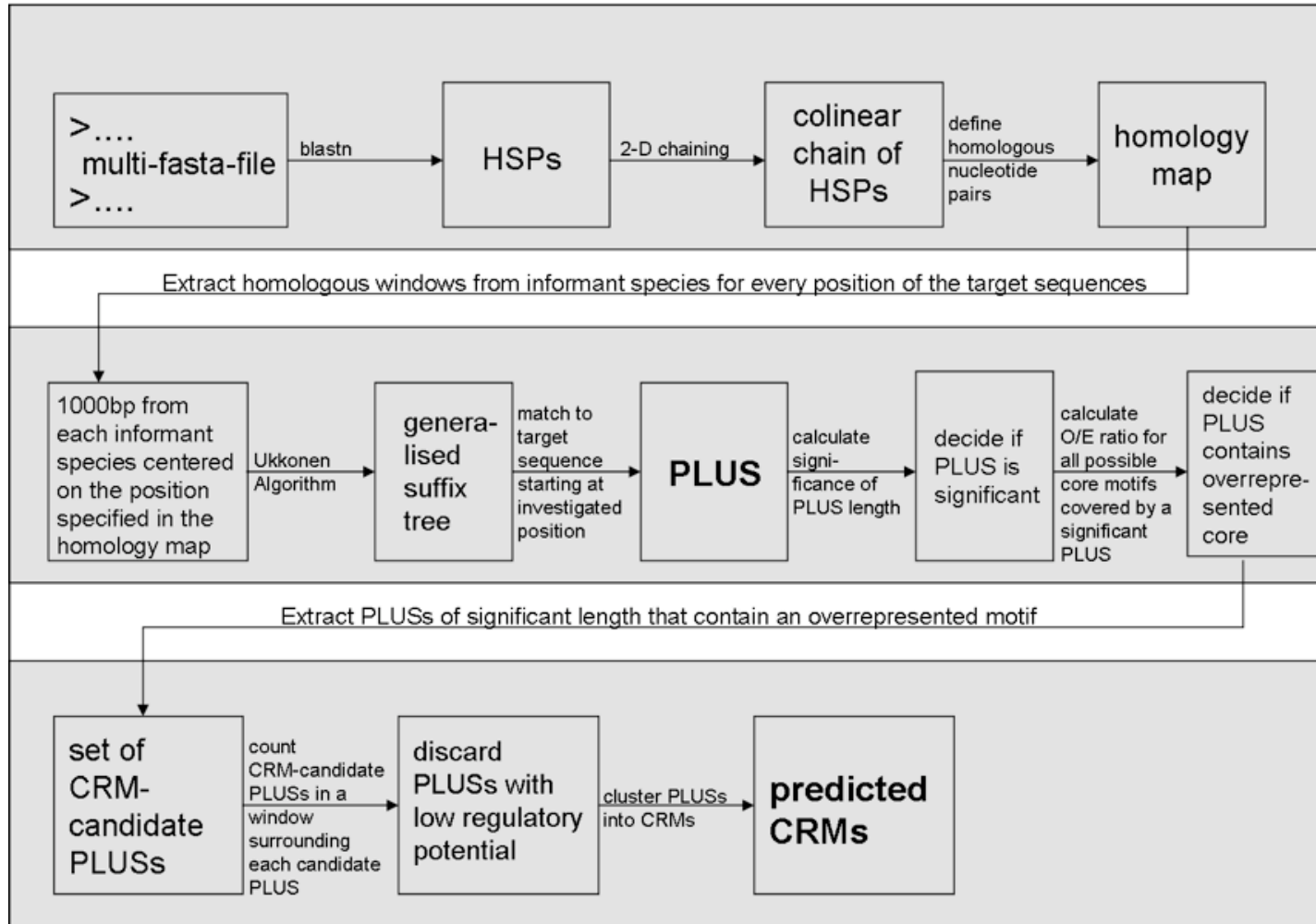


Motif  
↓  
Enhancer



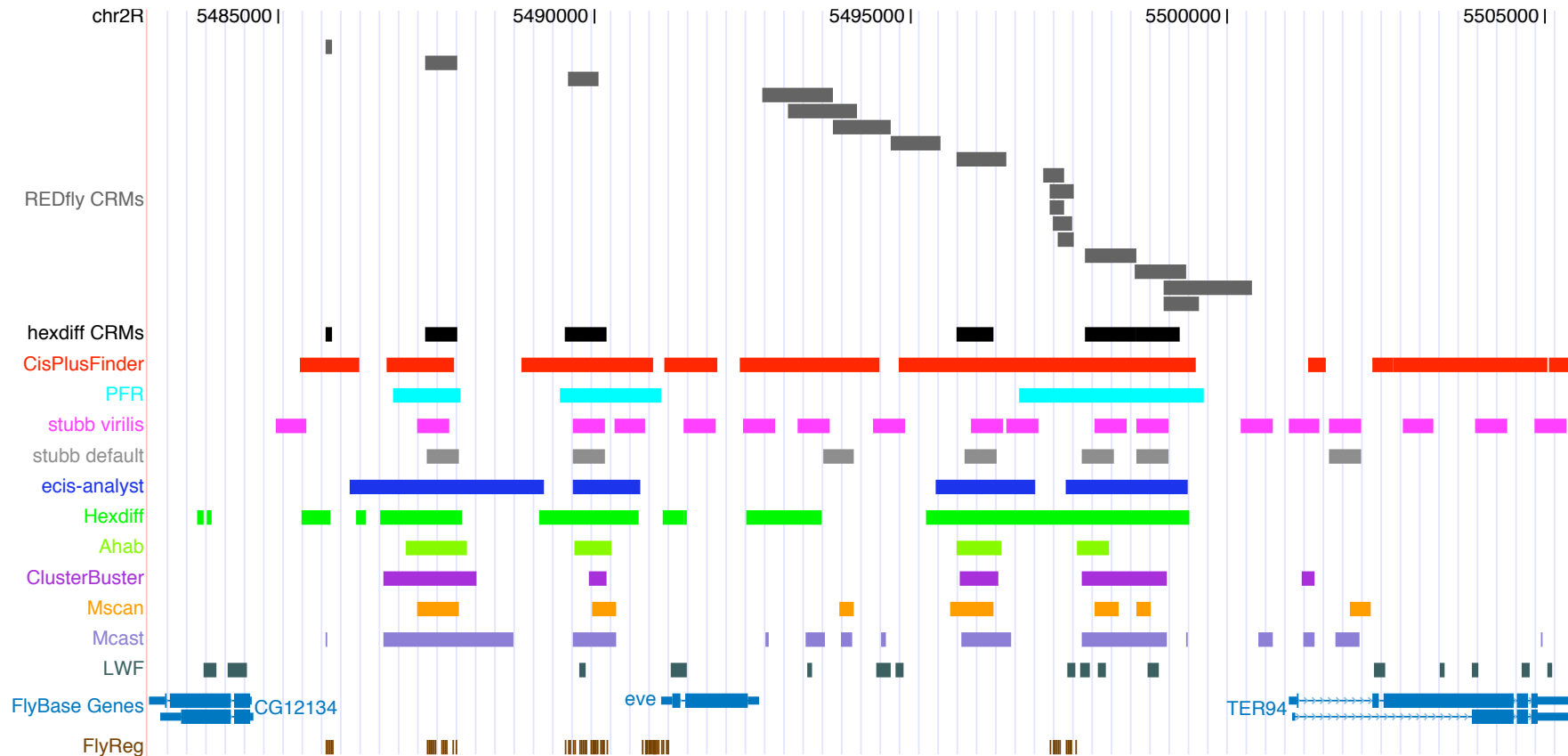
Enhancer  
↓  
Motif

# CisPlusFinder: a method to predict enhancers using conservation and over-representation



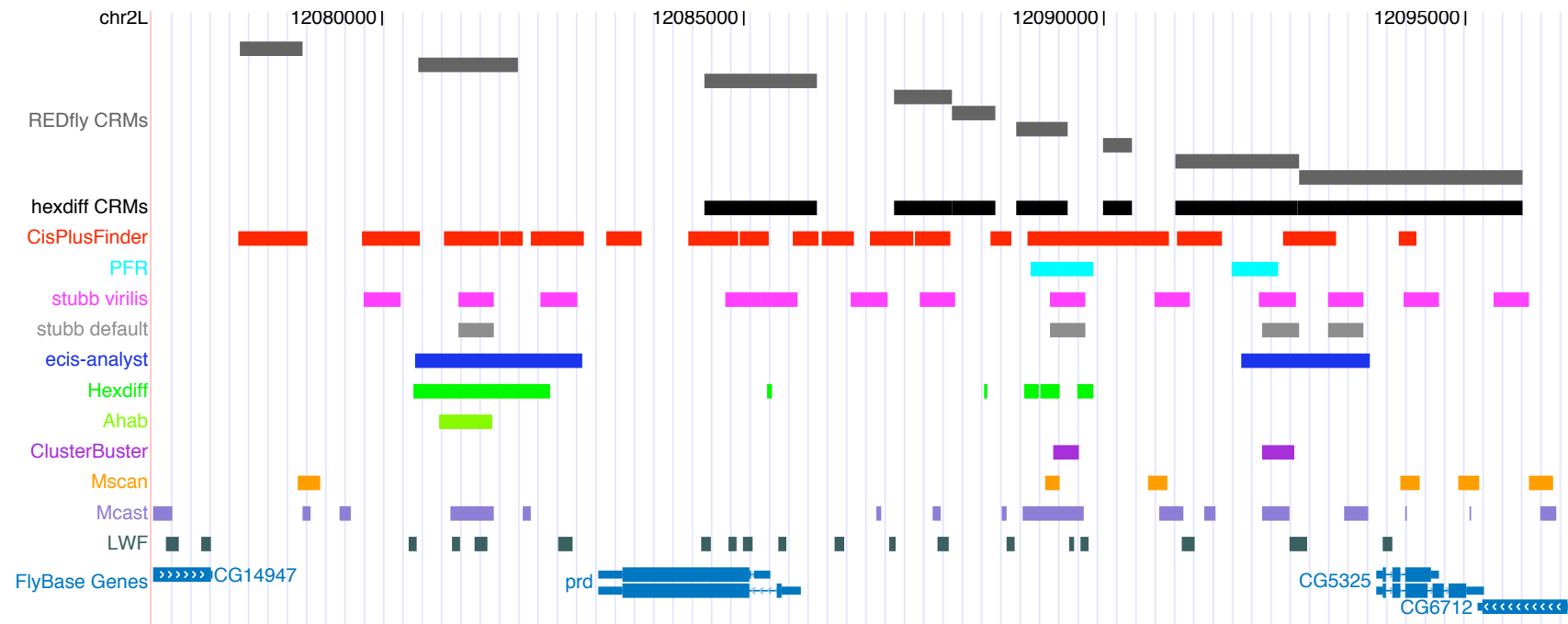
Pierstorff, Bergman & Wiehe (2006) *Bioinformatics* 22: 2858-2864

# Comparison of enhancer prediction methods for genes regulating *Drosophila* A-P development




Pierstorff, Bergman & Wiehe (2006) *Bioinformatics* 22: 2858-2864

# Comparison of enhancer prediction methods for genes regulating *Drosophila* A-P development



# Matching inferred motifs to functions & factors: a cautionary tale

## Tiffin: Motif Details

<b>Motif</b>	TIFDMEM0000040.1
<b>Species</b>	Drosophila melanogaster
<b>Collections</b>	<a href="#">120 motifs inferred from D. melanogaster promoter regions</a> <a href="#">D. melanogaster promoter motifs with significant correlations to embryonic gene expression patterns</a>
<b>Expression</b>	<a href="#">embryonic/larval fat body (P=0E0)</a> <a href="#">fat body (P=0E0)</a> <a href="#">amnioserosa anlage in statu nascendi (P=0E0)</a> <a href="#">fat body/gonad primordium (P=0E0)</a> <a href="#">extraembryonic structure (P=0E0)</a> <a href="#">amnioserosa (P=0E0)</a> <a href="#">fat body specific anlage (P=0E0)</a> <a href="#">adipose system (P=0E0)</a> <a href="#">embryonic/larval adipose system (P=0E0)</a>
<b>Weight matrix</b>	 (Unflip)
<b>Export</b>	<a href="#">[XMS]</a> <a href="#">[Frequencies]</a>



*serpent*



*pannier*

Down, Bergman, Su & Hubbard (2007) *PLoS Comp Biol* 3:e7.

# ORegAnno: Open Regulatory Annotation



AN OPEN ACCESS DATABASE FOR GENE REGULATORY ELEMENT AND POLYMORPHISM ANNOTATION

The Open REGulatory ANNOTation database (ORegAnno) is an open database for the curation of known regulatory elements from scientific literature. Annotation is collected from users worldwide for various biological assays and is automatically cross-referenced against PubMed, Entrez Gene, EnsEMBL, dbSNP, the eVOC: Cell type ontology, and the Taxonomy database, where appropriate, with information regarding the original experimentation performed (evidence). ORegAnno further provides an open validation process for all regulatory annotation in the public domain. Assigned validators receive notification of new records in the database and are able to cross-reference the citation to ensure record integrity. Validators have the ability to modify any record (deprecating the old record and creating a new one) if an error is found. Further, any contributor to the database can comment on any annotation by marking errors, or adding special reports into function as they see fit. These features of ORegAnno ensure that the collection is of the highest quality and uniquely provides a dynamic view of our changing understanding of gene regulation in the various genomes. As a first step, we recommend reading through our Help page.

The ORegAnno data and web application are all LGPL open-source to encourage the development and maintenance of the database to new information and experimentation techniques. Please use our current citation information when referring to ORegAnno data in publication. We encourage interested contributors to send email to the ORegAnno mailing list at [oreganno@bcgsc.ca](mailto:oreganno@bcgsc.ca) or to visit the [mailing-list archives](#).

## NEWS

June 26th, 2006 ORegAnno hits 500 unique publications annotated

June 20th, 2006 The RegCreative Jamboree (Nov. 29 to Dec. 1, 2006 in Ghent, Belgium) is now open for registration. [Register here](#).



April 18th, 2006 Introducing the ORegAnno Wiki Pages

April 11th, 2006 New version of ORegAnno released: Introducing the Publication Queue, a new way to suggest papers for annotation. Check it out by clicking on "Queue" in the menu to the right.

[More news...](#)

## MOST RECENTLY ANNOTATED PUBLICATIONS

Li LC et al., A single nucleotide polymorphism in the E-cadherin gene promoter alters transcriptional activities. *Cancer Res* 2000

Jormsjö S et al., Allele-specific regulation of matrix metalloproteinase-12 gene activity is associated with coronary artery luminal dimensions in diabetic patients with manifest coronary artery disease. *Circ Res* 2000

Ho Sui SJ et al., oPOSSUM: identification of over-represented transcription factor binding sites in co-expressed genes. *Nucleic Acids Res* 2005

Andrés V et al., Regulation of Gax homeobox gene transcription by a combination of positive factors including myocyte-specific enhancer factor 2. *Mol Cell Biol* 1995

Mutero A et al., Promoter elements of the mouse acetylcholinesterase gene. Transcriptional regulation during muscle differentiation. *J Biol Chem* 1995

You are not logged in

REGULATORY HAPLOTYPE: 7 entries  
REGULATORY REGION: 896 entries  
TRANSCRIPTION FACTOR BINDING SITE: 2071 entries  
REGULATORY POLYMORPHISM: 174 entries

[More details...](#)

menu



user menu



Montgomery et al. (2006) *Bioinformatics* 22:637-640

?



# Early ideas on the role of regulatory evolution

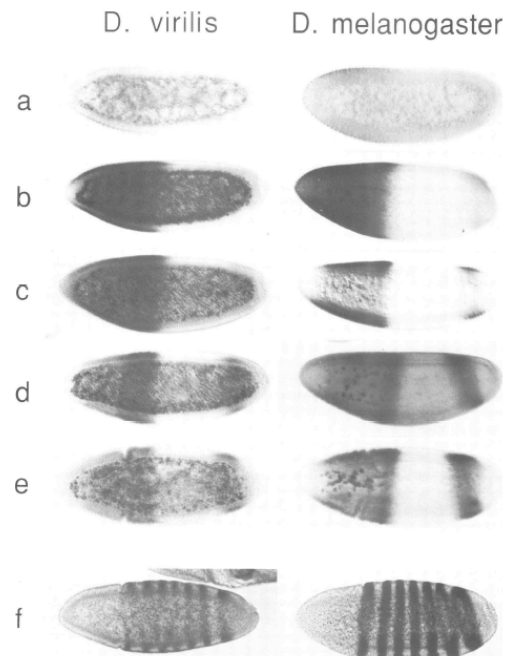
## DERIVATION OF CHAINS FROM PRECURSOR

Chains	Number of differences	Time (millions of years)	
$\alpha$ and $\beta$	78	-565	} Toward end of Pre-Cambrian
$\alpha$ and $\gamma$	~ 83	-600	
$\beta$ and $\gamma$	~ 36	-260	Beginning of Carboniferous
Horse and human	~ 18	(-130)	Cretaceous-Jurassic
$\beta$ and $\delta$	~ 6	- 44	Eocene
Gorilla $\alpha$	2	-14.5	} -11 Pliocene
and Human $\beta$	1	- 7.3	

Zukerkandl & Pauling (1963) - molecular clock

Wilson and colleagues (1970s) - discrepancy between rates of molecular and morphological evolution

# Widespread evidence for regulatory evolution in *Drosophila*



Treier et al. (1989)

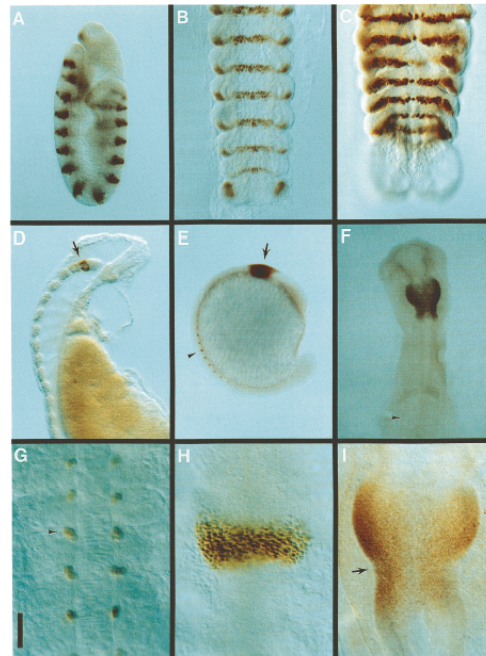
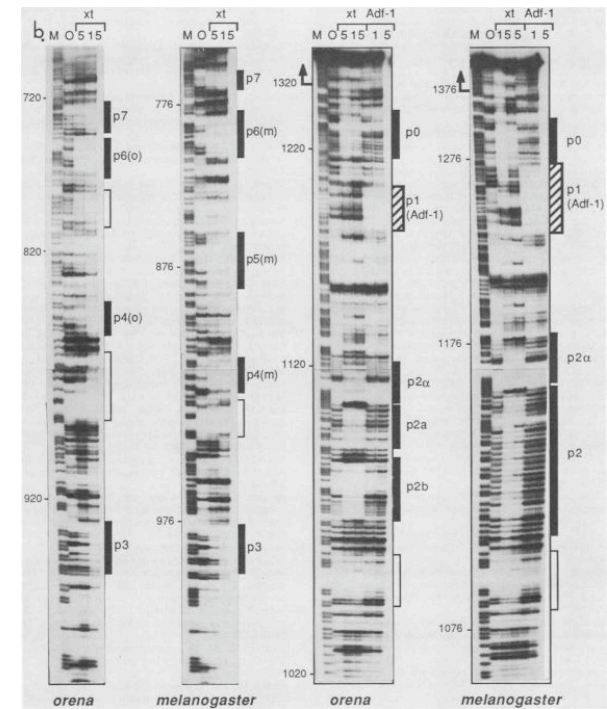


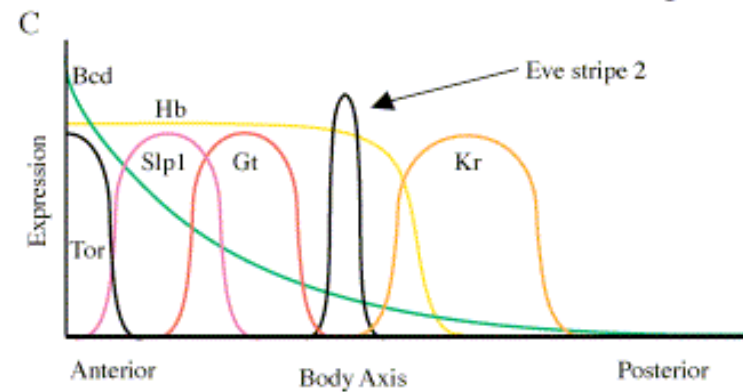
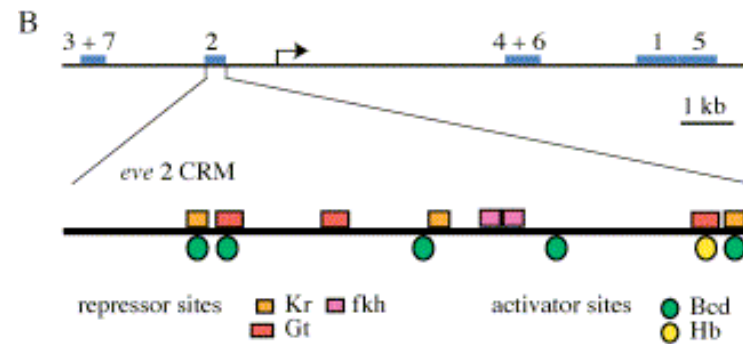
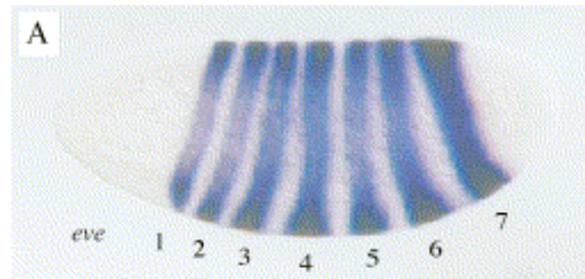
Figure 1. Expression of engrailed Proteins in Arthropods, Annelids, and Chordates  
Staining of *Urosalpinx* (A), grasshopper (B), crayfish (C), leech (D), zebrafish (E, G, H), and chick (F, I) embryos with the 4D8 MAb and HRP immunocytochemistry. In all three arthropods (A, B, C), engrailed is expressed in the posterior region of each segment. In the other organisms examined (D-I), engrailed does not appear to play a role during metameric development (for example, arrowhead in F marks the developing scrotes in the chick).

Patel et al. (1990)

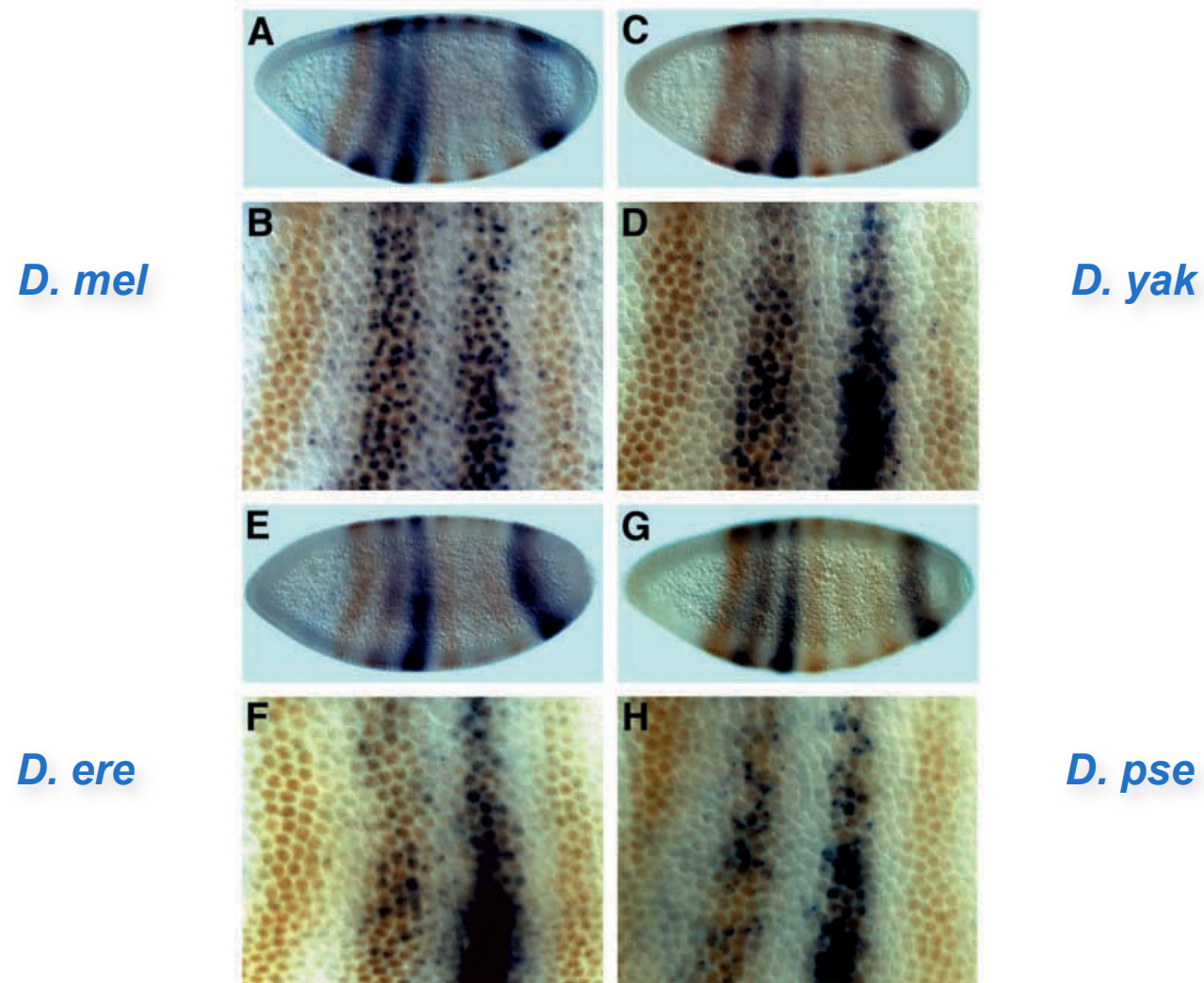


Moses et al. (1990)

# Regulation of *even-skipped* (*eve*)

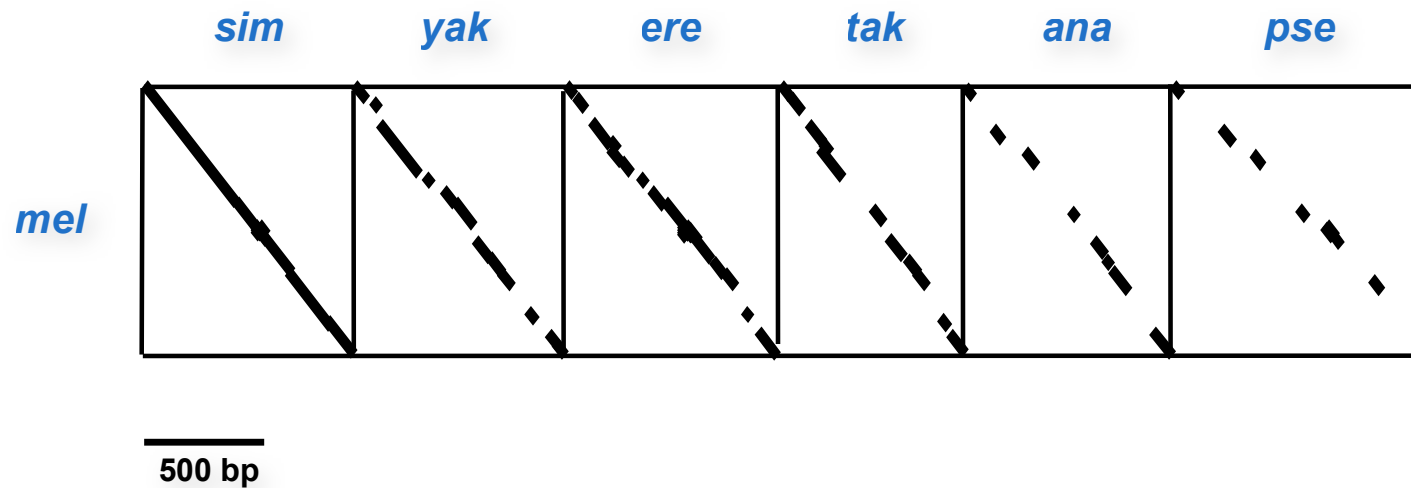


# Conservation of the eve stripe 2 enhancer pattern



Ludwig et al. (1998) *Development* 125:949-958

# Divergence in eve stripe 2 enhancer sequence



# Divergence in eve stripe 2 enhancer sequence

	kr-6	kr-5	kr-4	kr-3	kr-2	kr-1
<i>mel</i>	ATAACCCAAAT	TTAATCCGGT	ACC--GGGTTGC	GAAGGGATTAG	ACTGGGTAT	TTAACCCGTTT
<i>sim</i>	.....	.....	.....	.....	.....	.....
<i>yak</i>	.....	.....	.....	...C.....	.....	.....
<i>ere</i>	.....	.....	.....	...C.....	.....	.....
<i>pse</i>	.....	.....	..AA.....	.....A	.TC.....	.....C..G
<i>pic</i>	.....	.....	.....A.	AGG.....	.T...C....	C...G..AC.G

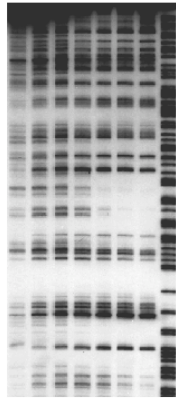
	bcd-5	bcd-4	bcd-3	bcd-2	bcd-1
<i>mel</i>	GTTAATCCG	GAGATTATT	TATAATCGC	GGGATTAGC	GAAGGGATTAG
<i>sim</i>	.....	C.....	.....	.....	.....
<i>yak</i>	.....	C.....	.C.....	.....	.....
<i>ere</i>	.....	C.....	.GC.C...G	.....	...C.....
<i>pse</i>	.....	C.....	...GT....	.....	...C.....
<i>pic</i>	.....	A.....	N/A	A.....	.....A
	.....	.C.....	N/A	.A.....G	AGG.....

	hb-3	hb-2	hb-1	gt-3
<i>mel</i>	CATAAAA-ACA	TTATTTTTT	CGATTTTTT	CGAGATTATTAGTCAATTG-----CAGITGC
<i>sim</i>	.....	.....	...C.....	.C.....
<i>yak</i>	.....	.....G	.T.C.....	.C.....
<i>ere</i>	.....	.....	..C.....	.C.....A.
<i>pse</i>	.....C...	C.....	N/A	.C.....C..
<i>pic</i>	.....C...G	.....C...	N/A	.C.....TTCATAITTC...C.-
	..C.C...G	.....C...	N/A	...C.....C..T...TTCC-AITT-.TC.CTA

	gt-2	gt-1
<i>mel</i>	GACITTTATTGCAGCATCTTG----AACAAATCGTC-GCAGTTTGGTAACAC	GAAAGTCATAAAA-ACACATAATA
<i>sim</i>	.....	.....
<i>yak</i>	.C.....G.....	.....
<i>ere</i>	.....CAGC.....G.....	.....
<i>pse</i>	..T.....AA.T.G.A.....T	.....C.....
<i>pic</i>	..T.....C.-T..AC.C.---.T.	.....C.C...G.....G

# Predicting binding site gains and losses

DNase I footprint



*bicoid*

Sequence data

ACGTAATCCCC

TGCTAAGCTGG

CGCTAAGCTCC

GGATCATCCAA

CTCTAATCCAG

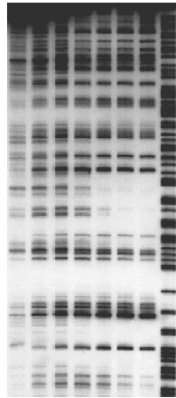
Position Weight Matrix (PWM)

	1	2	3	4	5	6	7	8
A	0.65	0.29	3.07	3.07	0.09	0.09	0.29	0.47
C	1.50	0.13	0.13	0.13	0.13	4.60	3.16	2.92
G	0.40	0.09	0.09	0.09	0.09	0.09	0.39	1.19
T	1.42	2.89	0.12	0.12	3.10	0.12	0.67	0.12

Predict new binding sites *in silico*

# Predicting binding site gains and losses

DNase I footprint

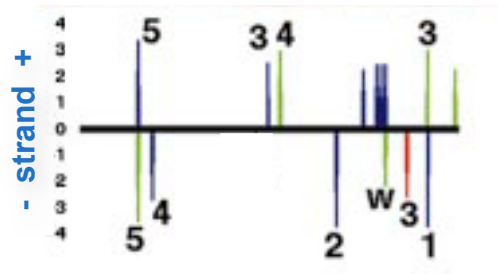


Sequence data

→ **ACGTAATCCCC**  
 → **TGCTAAGCTGG**  
 → **CGCTAAGCTCC**  
 → **GGATCATCCAA**  
 → **CTCTAATCCAG**

Position Weight Matrix (PWM)

	1	2	3	4	5	6	7	8
A	0.65	0.29	3.07	3.07	0.09	0.09	0.29	0.47
C	1.50	0.13	0.13	0.13	0.13	4.60	3.16	2.92
G	0.40	0.09	0.09	0.09	0.09	0.09	0.39	1.19
T	1.42	2.89	0.12	0.12	3.10	0.12	0.67	0.12



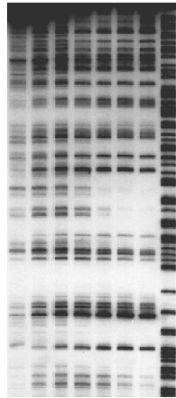
- *bicoid*
- *hunchback*
- *Kruppel*





# Predicting binding site gains and losses

DNase I footprint

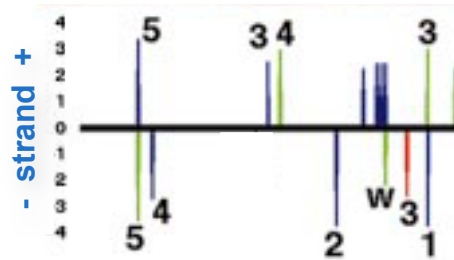


Sequence data

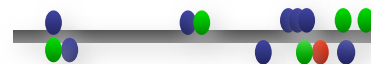
→ **ACGTAATCCCC**  
 → **TGCTAAGCTGG**  
 → **CGCTAAGCTCC**  
 → **GGATCATCCAA**  
 → **CTCTAATCCAG**

Position Weight Matrix (PWM)

	1	2	3	4	5	6	7	8
A	0.65	0.29	3.07	3.07	0.09	0.09	0.29	0.47
C	1.50	0.13	0.13	0.13	0.13	4.60	3.16	2.92
G	0.40	0.09	0.09	0.09	0.09	0.09	0.39	1.19
T	1.42	2.89	0.12	0.12	3.10	0.12	0.67	0.12



- *bicoid*
- *hunchback*
- *Kruppel*



*D. melanogaster*

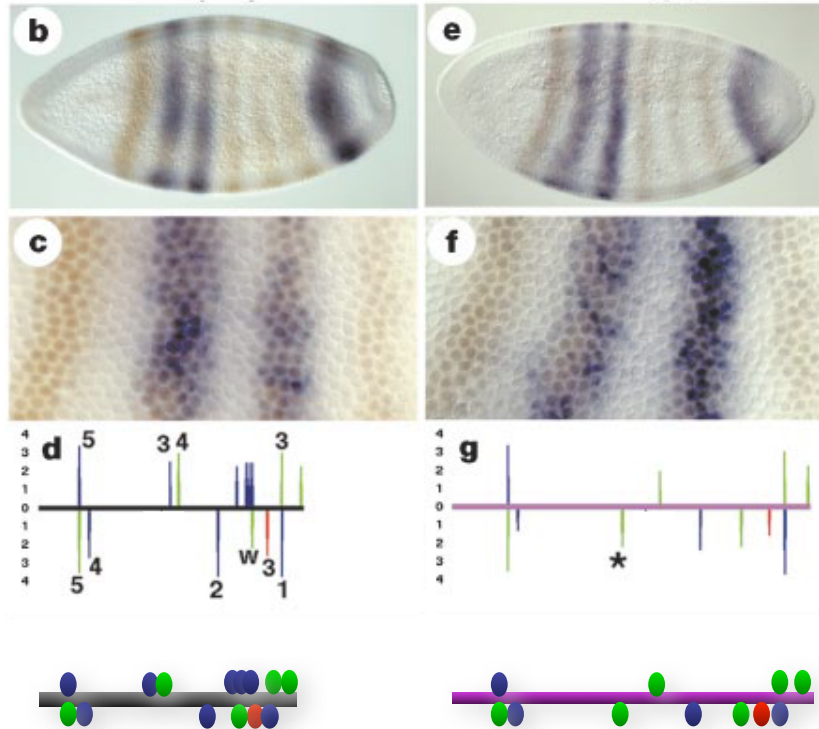


*D. pseudoobscura*

# A model for enhancer evolution: binding site turnover under stabilizing selection

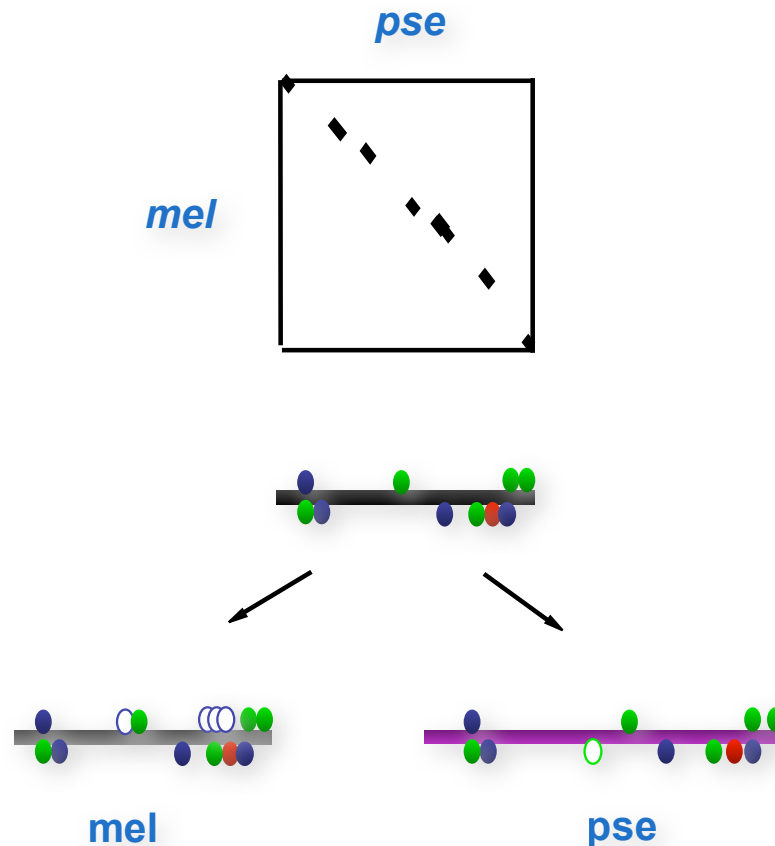
*D. mel*  
in *D. mel*

*D. pseudo*  
in *D. mel*

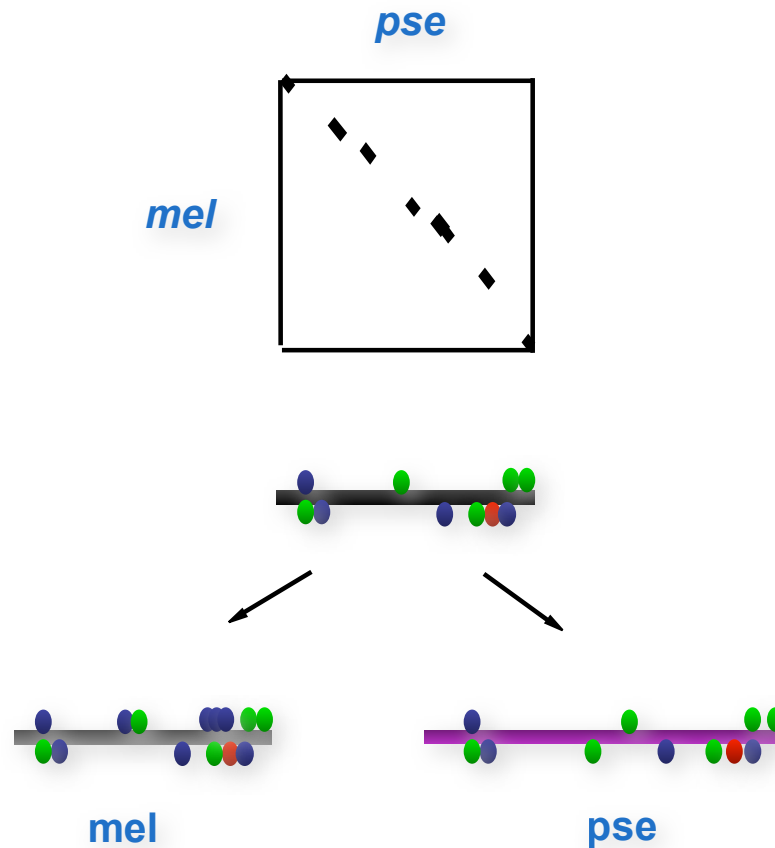


Ludwig et al. (200) *Nature* 403:564-567

**Is eve stripe 2 divergence strictly neutral:  
 $H_0$ : are ancestral sites sufficient for function?**



**Is eve stripe 2 divergence strictly neutral:  
 $H_a$ : do derived sites contribute to function?**

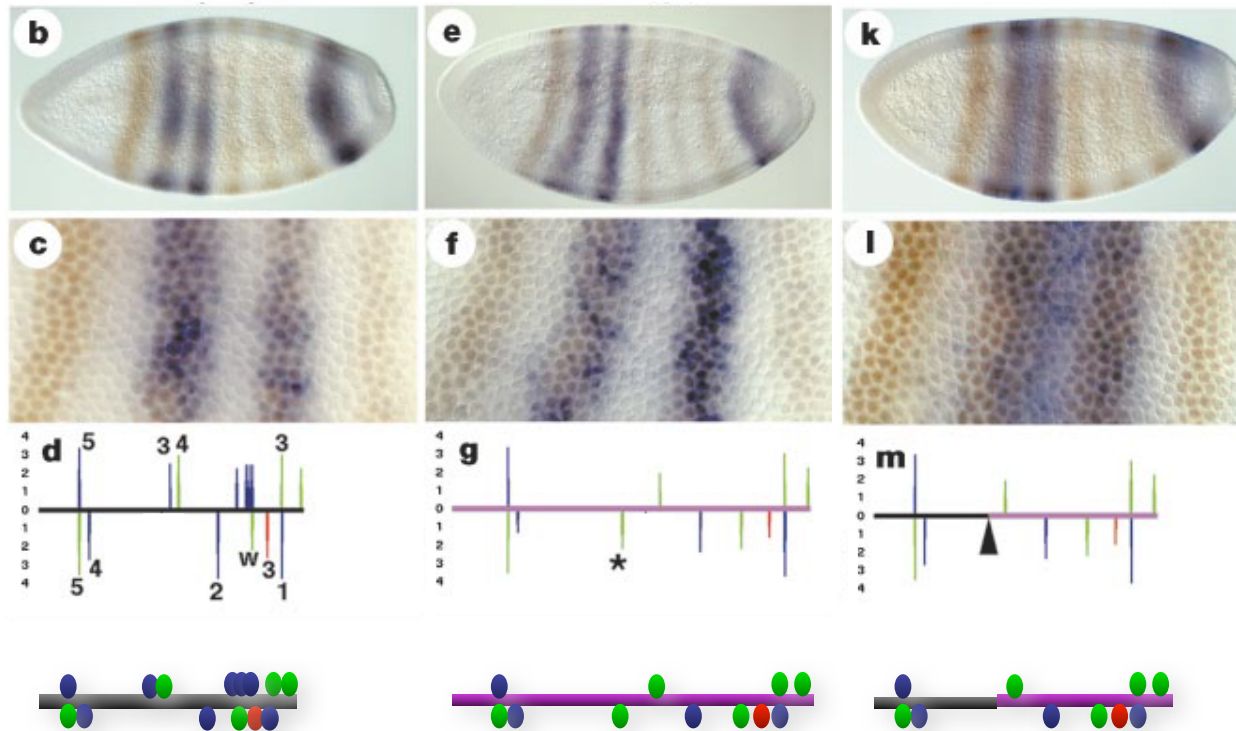


# A model for enhancer evolution: binding site turnover under stabilizing selection

*D. mel*  
in *D. mel*

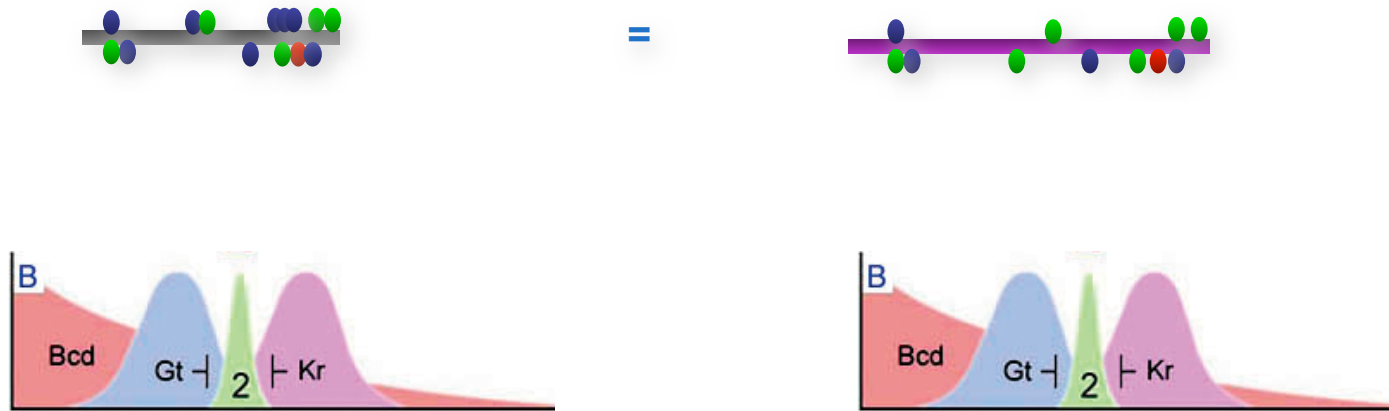
*D. pseudo*  
in *D. mel*

*mel-pseudo chimera*  
in *D. mel*

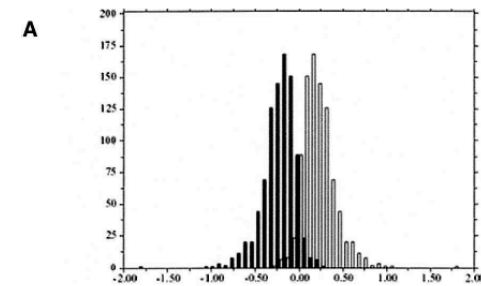
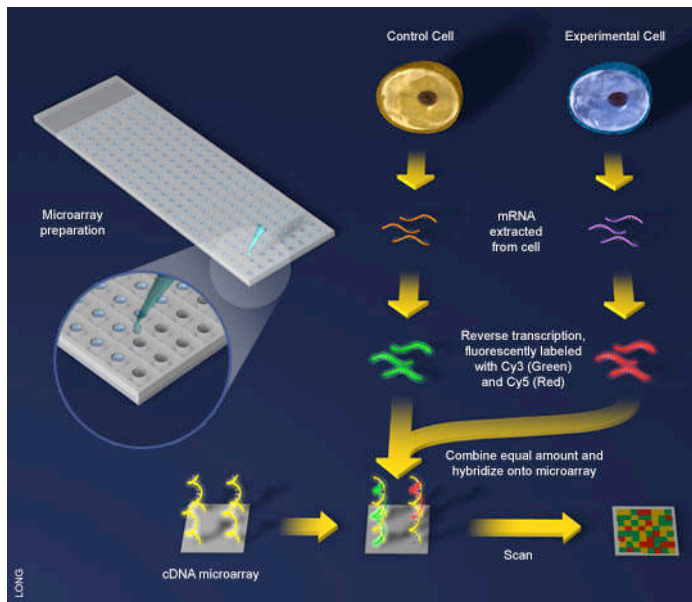


Ludwig et al. (200) *Nature* 403:564-567

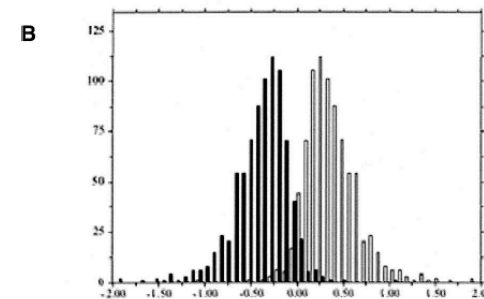
# Is stabilizing selection on *cis*-regulatory sequences a paradigm for gene expression evolution?



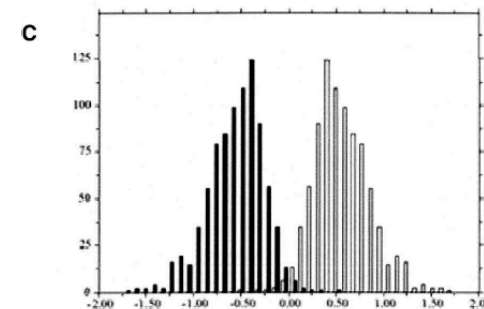
# Microarrays & gene expression evolution



h-c



h-o



h-m

Gilad et al (2005) Gen. Res.

# Extent of regulatory evolution in *Drosophila*

Larval-pupal transition - 4 strains *D. mel* + *D. sim* + *D. yak*

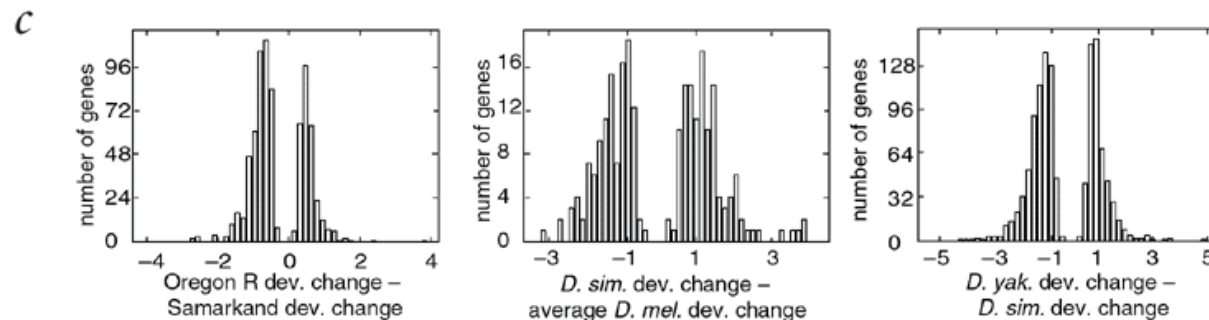
cDNA - 13,000 genes - 50% developmentally regulated

27% of genes show change between 2 lineages

Table 1 • Distances between lineages<sup>a</sup>

	Canton S	Oregon R	Netherlands2	<i>D. simulans</i>	<i>D. yakuba</i>
Samarkand	355	753	217	779	1,472
Canton S		549	241	629	1,023
Oregon R			807	827	1,348
Netherlands2				754	1,286
<i>D. simulans</i>					1,162

<sup>a</sup>Shown are the numbers of genes with significantly different changes in developmental expression between two lineages. For each pair-wise comparison, we considered only genes that change developmentally in either lineage: 3,457 distinct genes differ in developmental changes between at least two lineages.



Rifkin et al (2003) Nat. Genet.

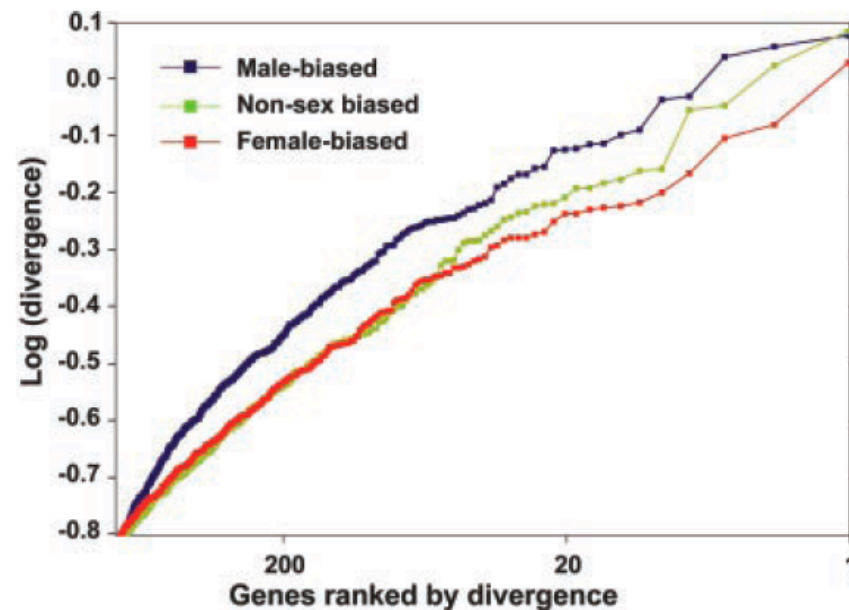


# Extent of regulatory evolution in *Drosophila*

Adult males and females - 8 strains *D. mel* + *D. sim*

cDNA - 5,000 genes - 58% sex biased

50% vary between species, 40% vary between lines



Ranz et al (2003) Science

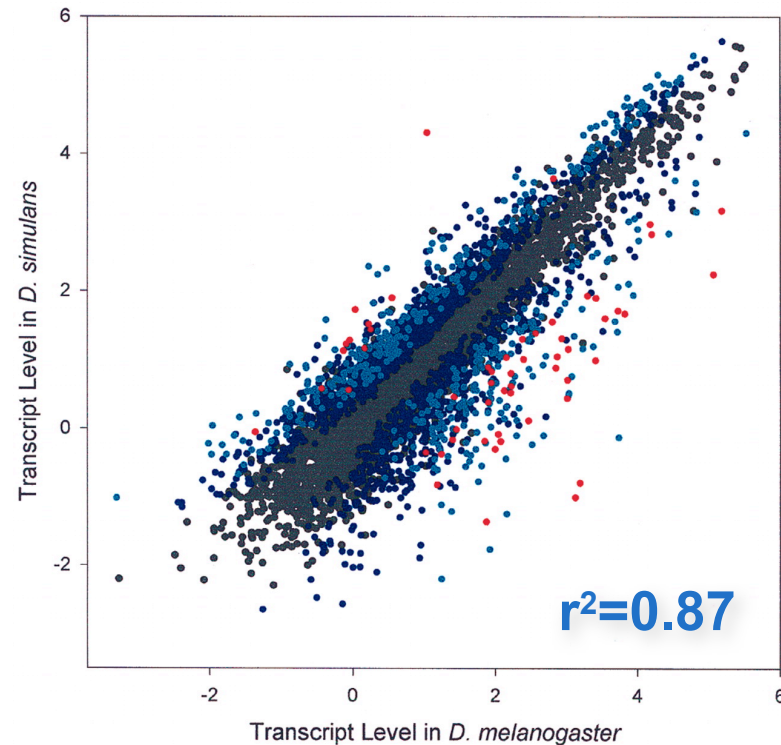
Mieklejohn et al (2003) PNAS

# Extent of regulatory evolution in *Drosophila*

Adult males - 10 strains *D. sim* + 10-strain pool of *D. mel*

Affymetrix - 14,000 'genes' - 6,700 in both species

34% vary between species, 21% vary between lines



Nuzhdin et al (2004) Mol Biol Evol

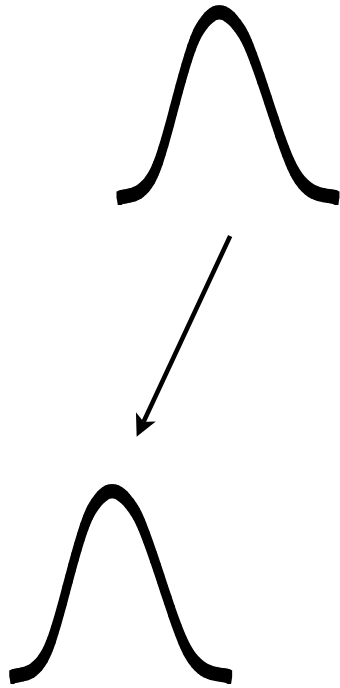
# Variation in levels of gene expression is common within species

Organism	Citation	Source of error variance	Statistical criteria	Proportion of genes variable within taxa	Proportion of genes variable among populations
<b>Yeast</b>					
Among progeny within strain	Cavalieri <i>et al.</i> (2000)	No error variance	Twofold difference	6%	
Among isolates	Townsend <i>et al.</i> (2003)	5 replicate slides (dye swap)	95% credible intervals, Bayesian analysis	7%	
Among inbred lines	Brem <i>et al.</i> (2002)	6 replicate cultures	$P < 0.005$ , Wilcoxon-Mann-Whitney test	25%	
<b>Fruit flies</b>					
Among inbred strains	Jin <i>et al.</i> (2001)	6 replicate slides (dye swap), pooled samples	nonzero contribution to variance	~25%	
<b>Maize</b>					
Among inbred strains	Schadt <i>et al.</i> (2003)	dye swap	$P < 0.05$ for at least 10% of samples	77%	
<b>Mice</b>					
Within strain	Pritchard <i>et al.</i> (2001)	4 replicate slides (dye swap)	$P < 0.05$ for ANOVA	3.3%	
Among offspring or parents of cross	Schadt <i>et al.</i> (2003)	dye swap	$P < 0.05$ for at least 10% of samples	33%	
Among inbred strains	Sandberg <i>et al.</i> (2000)	2 replicate mice from each strain	$> 1.8$ -fold change	1%	
<b>Humans</b>					
Among related individuals	Cheung <i>et al.</i> (2003)	4 replicate slides	$s^2$ among individuals $> s^2$ among replicates	Not determined	
Among unrelated individuals	Cobb <i>et al.</i> (2005)	Not reported	$s^2$ among individuals $> s^2$ among replicates	Not determined	
	Morley <i>et al.</i> (2004)	2 replicate slides	$s^2$ among individuals $> s^2$ among replicates	42%	
	Whitney <i>et al.</i> (2003)	Not reported	Reference design, $>$ twofold change	Not determined	
<b>Fish</b>					
Among unrelated individuals	Oleksiak <i>et al.</i> (2002)	4 replicate slides (dye swap) + 2 spots/slide	$P < 0.05$ for ANOVA	18%	2%
	Oleksiak <i>et al.</i> (2005)	2 replicate slides (dye swap) + 8 spots/slide	$P < 0.01$ for ANOVA	94%	10%
	Whitehead & Crawford (2005b)	4 replicate slides (dye swap) + 6 spots/slide	$P < 0.05$ for ANOVA*	82%	3%
	Whitehead & Crawford (2005a)	2 replicate slides (dye swap) + 6 spots/slide	$P < 0.05$ for ANOVA	69%	12%†

\*Data from Whitehead & Crawford (2005b) re-analysed using nested ANOVA design for purposes of comparison.

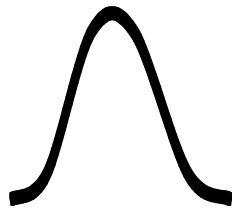
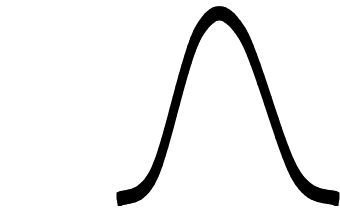
†Unpublished data from the study reported in Whitehead & Crawford (2005a).

# Alternative modes of gene expression divergence as a quantitative trait



mutation-drift  
equilibrium

# Alternative modes of gene expression divergence as a quantitative trait

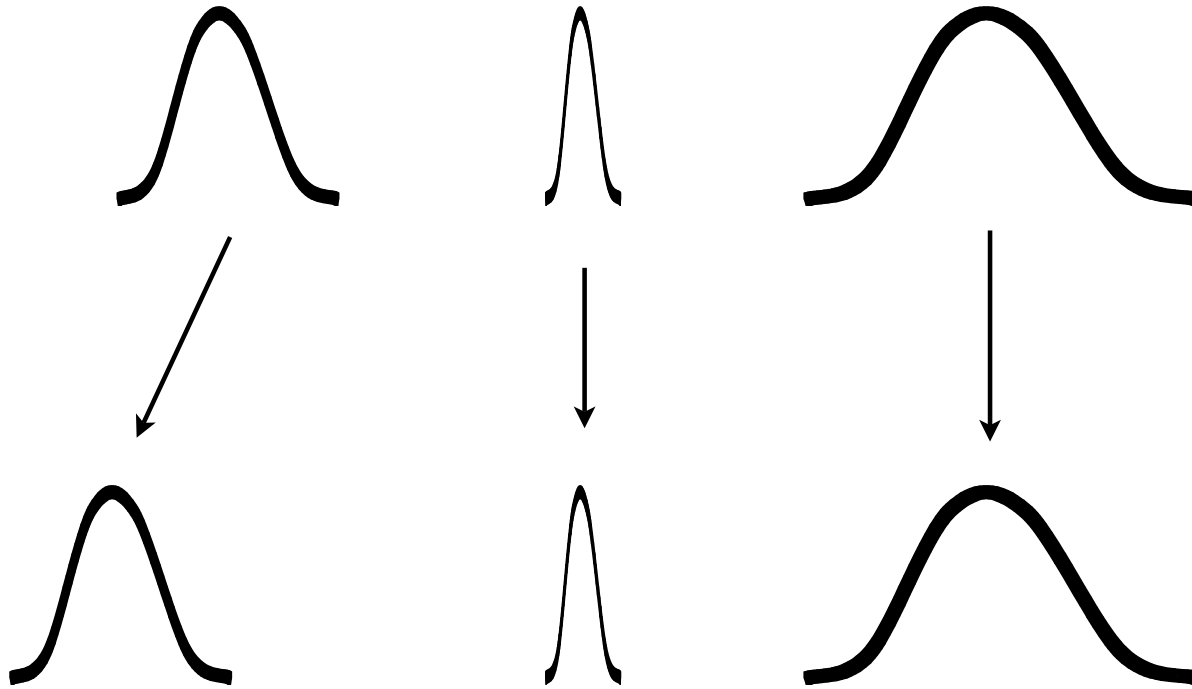


**mutation-drift  
equilibrium**



**stabilizing  
selection**  
or low  
mutation rate

# Alternative modes of gene expression divergence as a quantitative trait

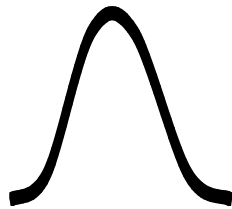
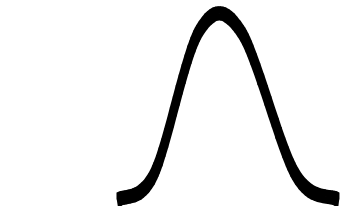


mutation-drift  
equilibrium

stabilizing  
selection  
or low  
mutation rate

balancing  
selection

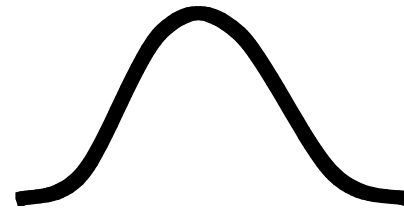
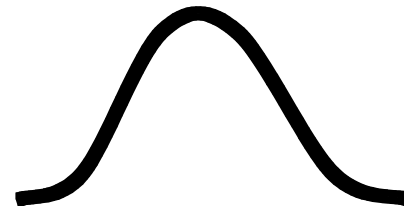
# Alternative modes of gene expression divergence as a quantitative trait



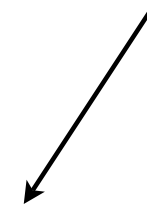
**mutation-drift  
equilibrium**



**stabilizing  
selection**  
or low  
mutation rate

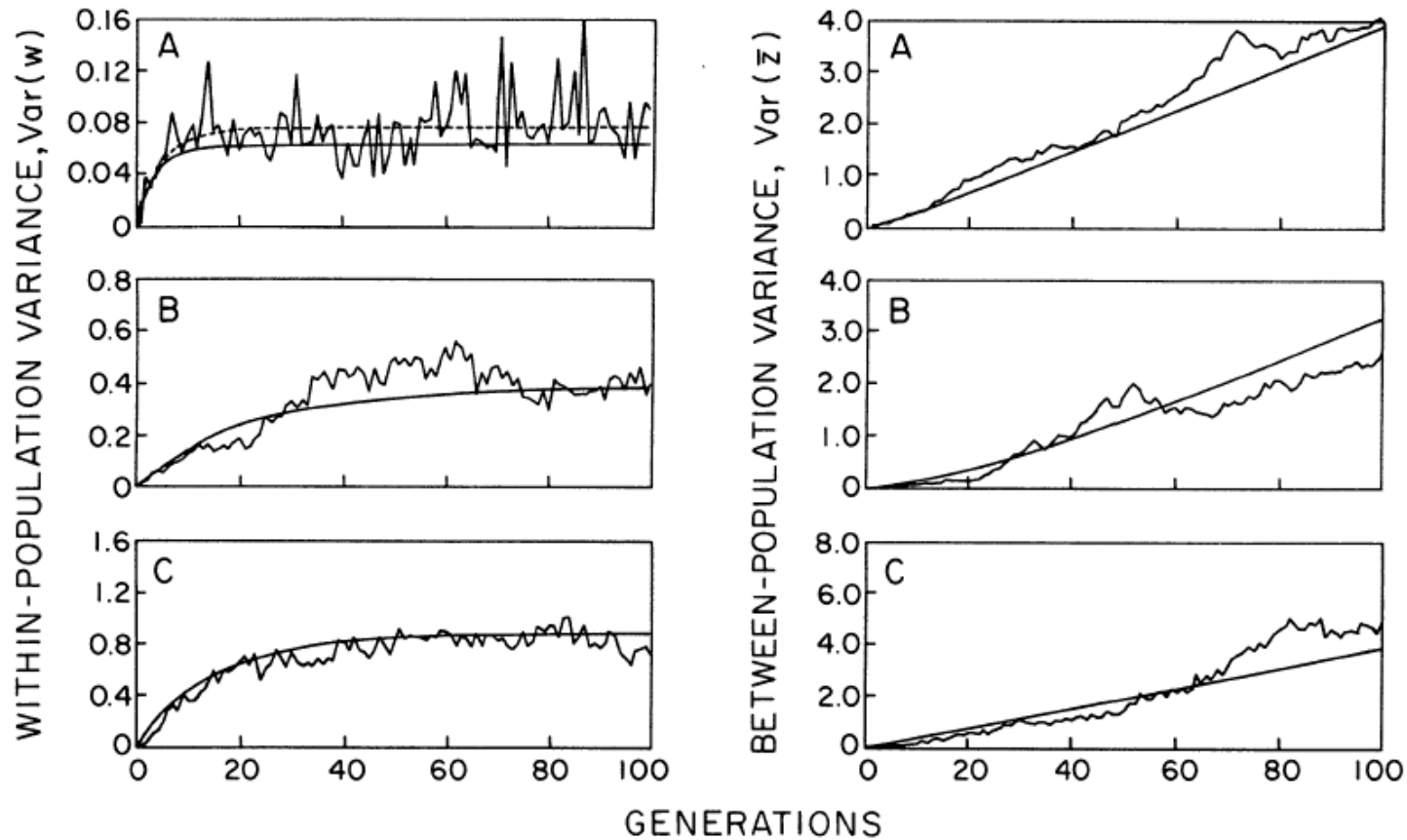


**balancing  
selection**



**directional  
selection**

# Mutation-drift equilibrium (MDE) model of evolution for a quantitative trait



$$E(V_w) \approx 2 N_e * V_m$$

$$E(V_b) \approx 2 V_m * t$$

Lynch & Hill (1986) Evolution



# Testing mutation-drift equilibrium (MDE) for a quantitative character I

$$F = \frac{V_b}{V_m * t}$$

Test null hypothesis using F-distribution

$V_b$  = variance between species

$V_m$  = mutational variance

$t$  = number of generations

# Testing mutation-drift equilibrium (MDE) for a quantitative character II

$$\Delta = \frac{V_b}{V_w * t}$$

Compare  $\Delta$  to “mutational heritability” =  $V_m / V_e$

$V_b$  = variance between species

$V_w$  = variance within species

$V_m$  = mutational variance

$V_e$  = environmental variance

t = number of generations

# Mode of regulatory evolution in primates, mice and flies

TABLE 1. Datasets analyzed in this study.

Contrast	Estimate of minimum divergence time	Minimum number of generations/ year	Estimate of minimum total number of generations $\times 10^3$	Dataset
Laboratory strains of <i>Mus musculus</i> (129SvEv $\times$ C57BL/6) 4 brain tissues	60 years	5	0.6	Sandberg et al. (2000)
<i>Drosophila melanogaster</i> African $\times$ North American strains	300 years	8	4.8	Meiklejohn et al. (2003)
Human $\times$ chimpanzee (primate liver)	6 million years	0.05	600	Enard et al. (2002)
Human $\times$ chimpanzee (primate brain)	6 million years	0.05	600	Enard et al. (2002)
<i>M. musculus</i> $\times$ <i>M. spretus</i> (brain)	2 million years	2	8000	Enard et al. (2002)
<i>M. musculus</i> $\times$ <i>M. spretus</i> (liver)	2 million years	2	8000	Enard et al. (2002)
<i>D. melanogaster</i> $\times$ <i>D. simulans</i>	2 million years	8	32,000	Ranz et al. (2003)

Assumed  $10^{-4} < \text{mutational heritability} < 10^{-2}$

$$\Delta = \frac{V_b}{V_w * t}$$

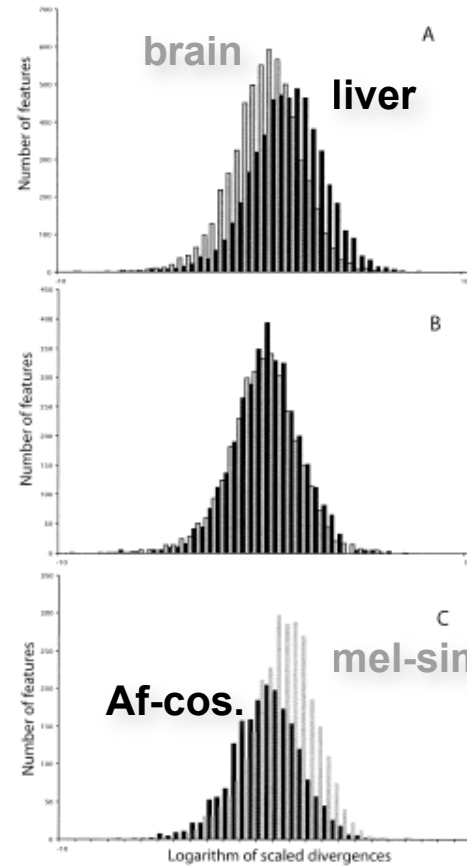
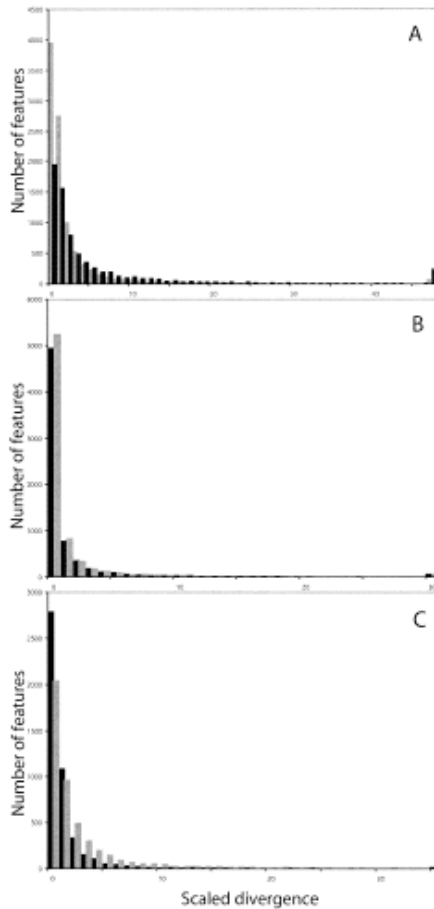
$10^{-4} < \Delta < 10^{-2}$  : drift

$\Delta > 10^{-2}$  : directional

$\Delta < 10^{-4}$  : stabilizing

# Mode of regulatory evolution in primates, mice and flies

$$\Delta t = \frac{V_b}{V_w}$$



primates

mice

flies

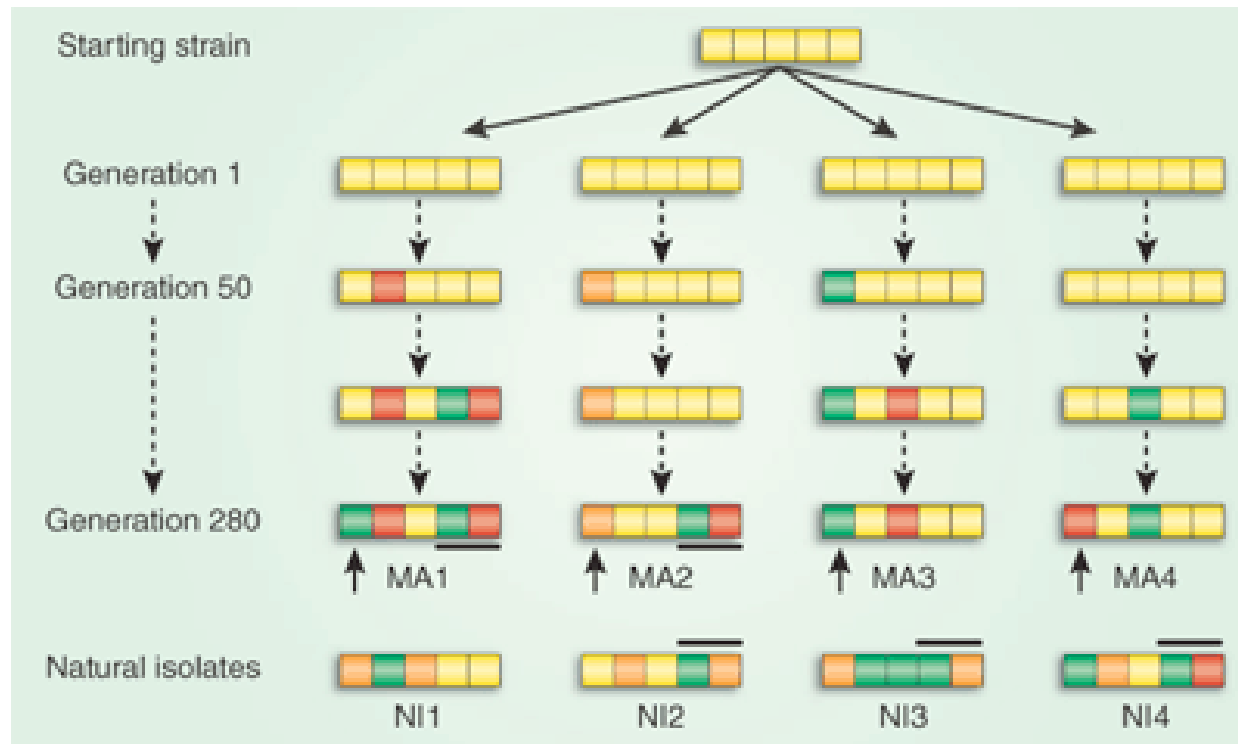
Lemos et al (2005) Evolution

# Mode of regulatory evolution in primates, mice and flies

TABLE 2. Number of features classified by rate of divergence. To be conservative and underestimate the number of features under stabilizing selection, we used lower estimates of the number generations separating the samples (see Table 1) and the unidirectional divergence (i.e., half the minimum number of generations).

Contrast	Stabilizing selection (rate < 0.0001)	Genetic drift/ mutation accumulation (0.0001 < rate < 0.01)	Directional selection (rate > 0.01)	Number of generations required for the top decile of rates to exceed 0.01
Mouse strains: hippocampus	6030 (67%)	2189 (24%)	823 (9%)	—
Mouse strains: cortex	5335 (61%)	2366 (27%)	1106 (12%)	—
Mouse strains: midbrain	6026 (66%)	2184 (24%)	885 (10%)	—
Mouse strains: cerebellum	5657 (63%)	2243 (25%)	1002 (12%)	—
<i>Drosophila melanogaster</i> : North American vs. African strains	4437 (93%)	306 (6%)	16 (1%)	230
Primates: liver	8761 (96%)	372 (4%)	0	1250
Primates: brain	9231 (99%)	95 (1%)	0	460
Mouse species: liver	6833 (100%)	0	0	300
Mouse species: brain	7252 (100%)	1 (<1%)	0	300
<i>D. melanogaster</i> vs. <i>D. simulans</i>	4758 (99%)	1 (1%)	0	650

# Mutational heritability in gene expression is the key (unknown) parameter

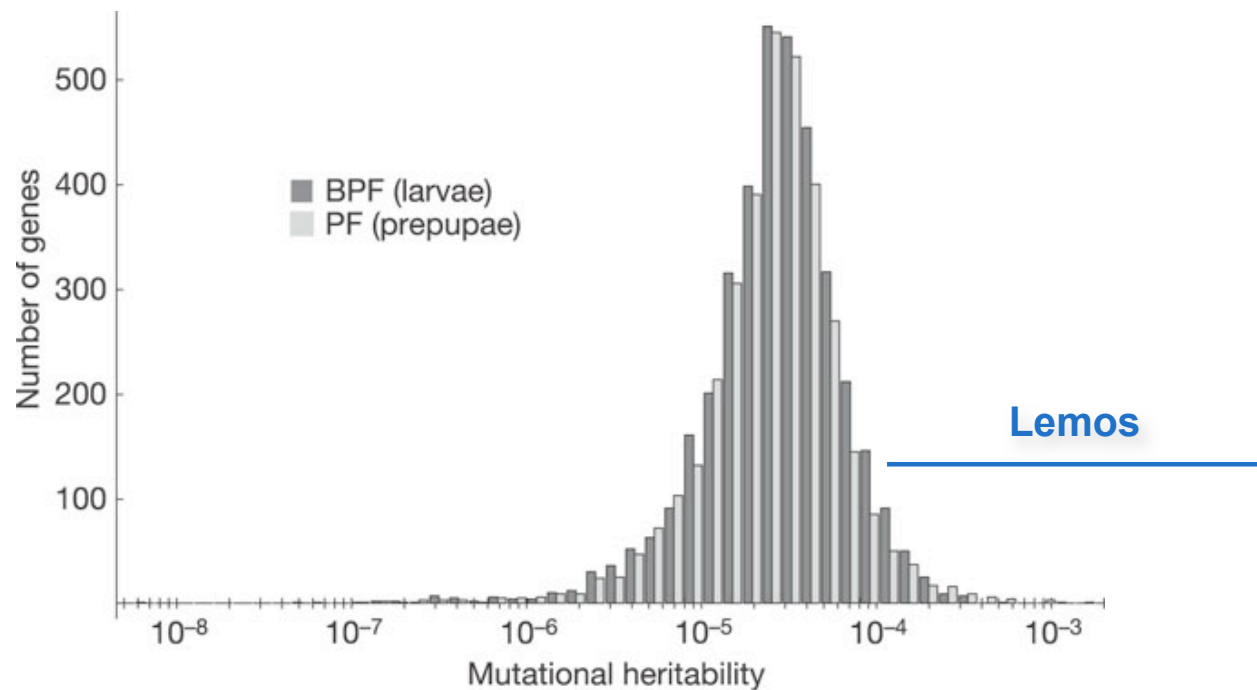


# Measuring the mutational variance of gene expression in flies

12 mutation accumulation lines - 200 generations

cDNA - 13,000 genes - larval-pupal transition

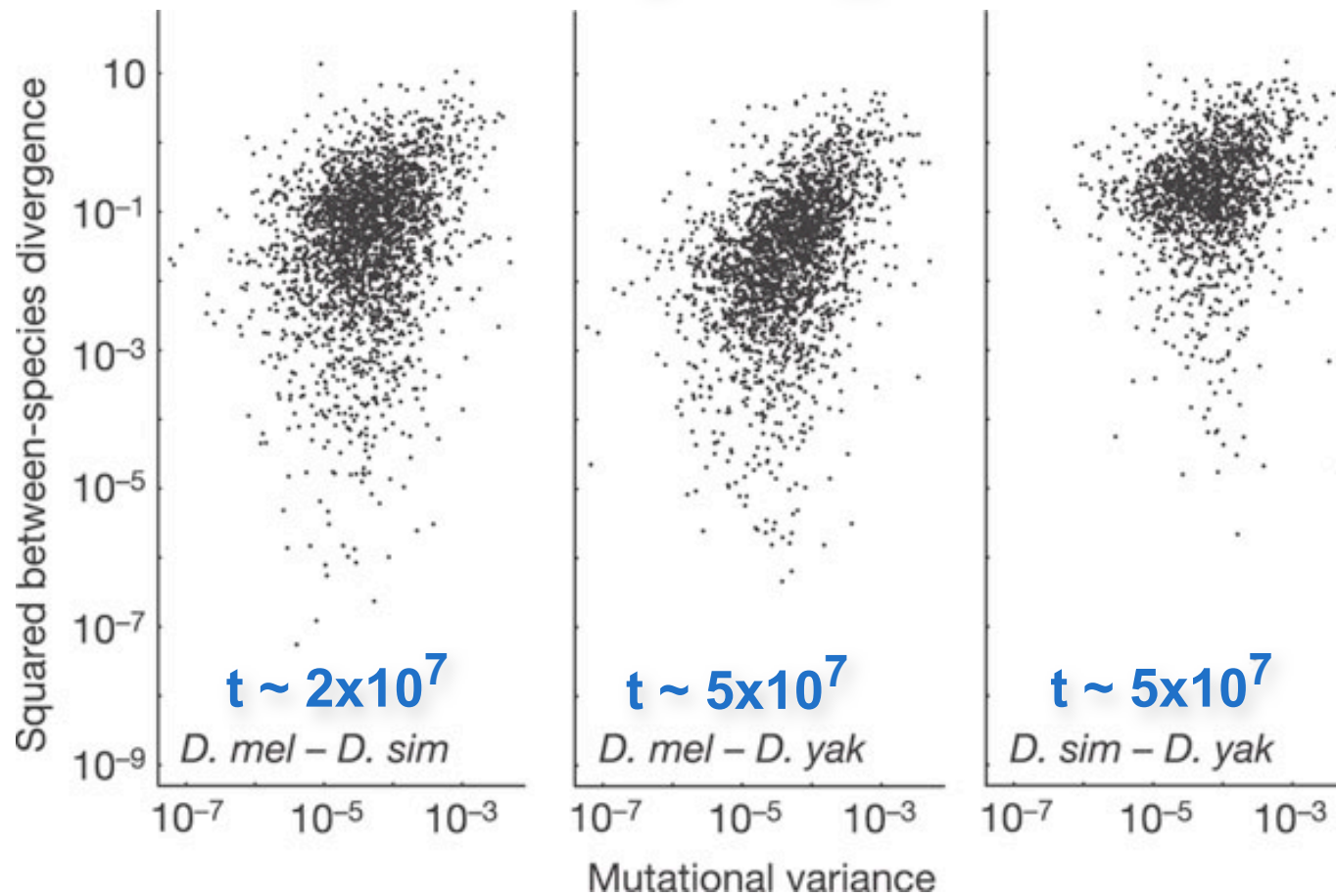
40% of genes show variation between MA lines



Rifkin et al (2005) Nature

# Measuring the mutational variance of gene expression in flies

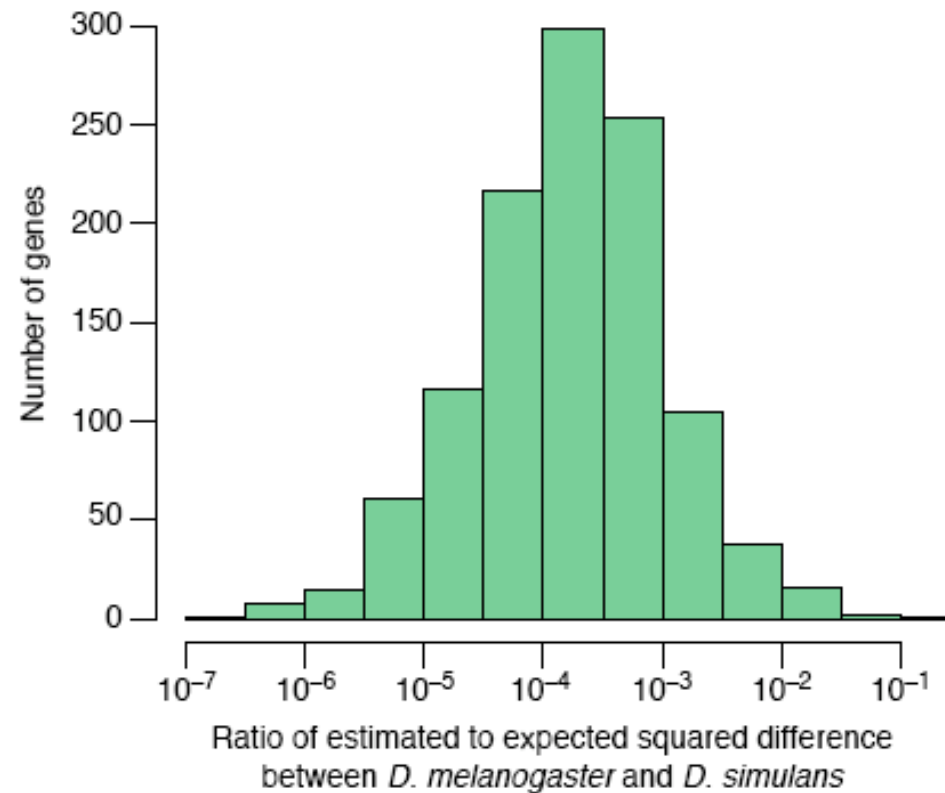
$$E(V_b) \approx 2 V_m * t$$



Rifkin et al (2005) Nature



# Stabilizing selection is the major mode of gene expression evolution in *Drosophila*



Gilad et al (2006) Trends Genet.

?

# Acknowledgements



**Marty Kreitman, Michael Ludwig,  
Nipam Patel**



**Sue Celniker, Gerry Rubin,  
Eddy Rubin**



**Michael Ashburner**

**Thomas Down - Wellcome Trust Sanger Institute**



**MARIE CURIE ACTIONS**

**Nora Pierstorff - Cologne**