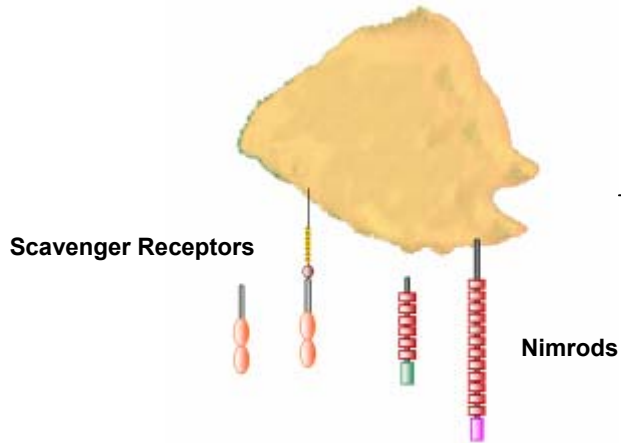


Evolution of the innate immunity gene regulatory network in *Drosophila*

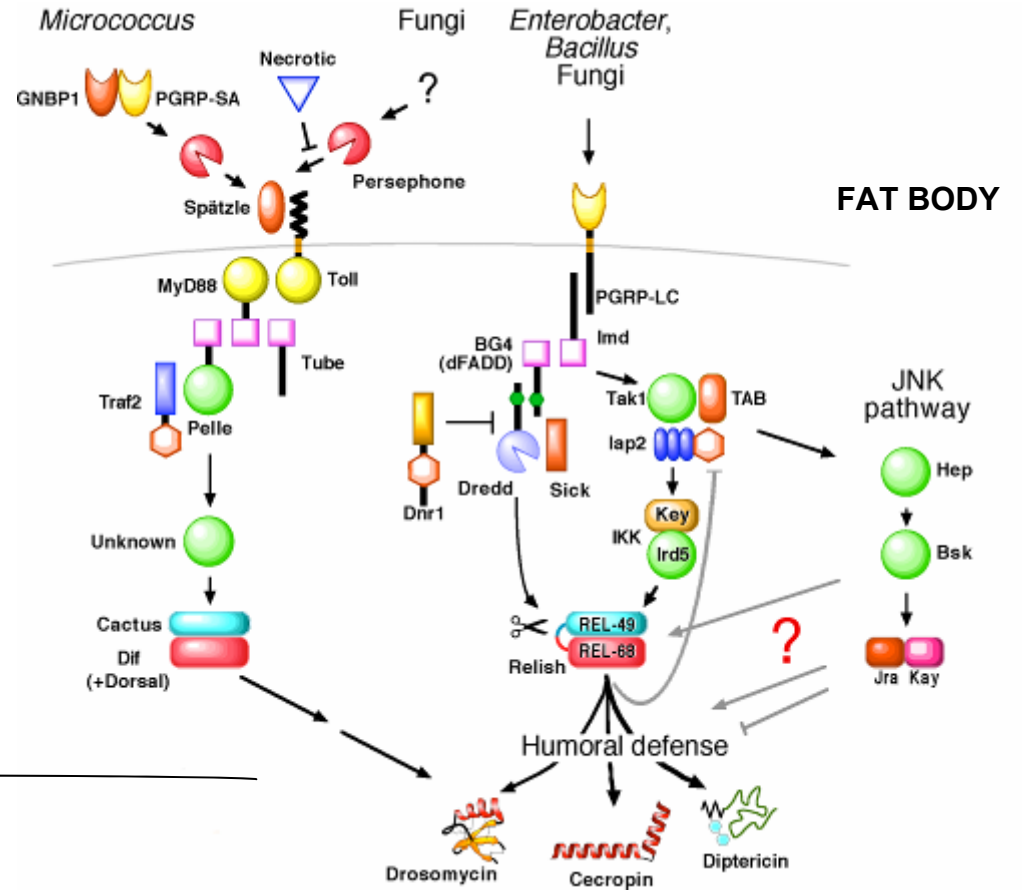
Andrew G. Clark
Cornell University
Oxford, March 3, 2007

What is the genetic basis for natural variability in immune competence?

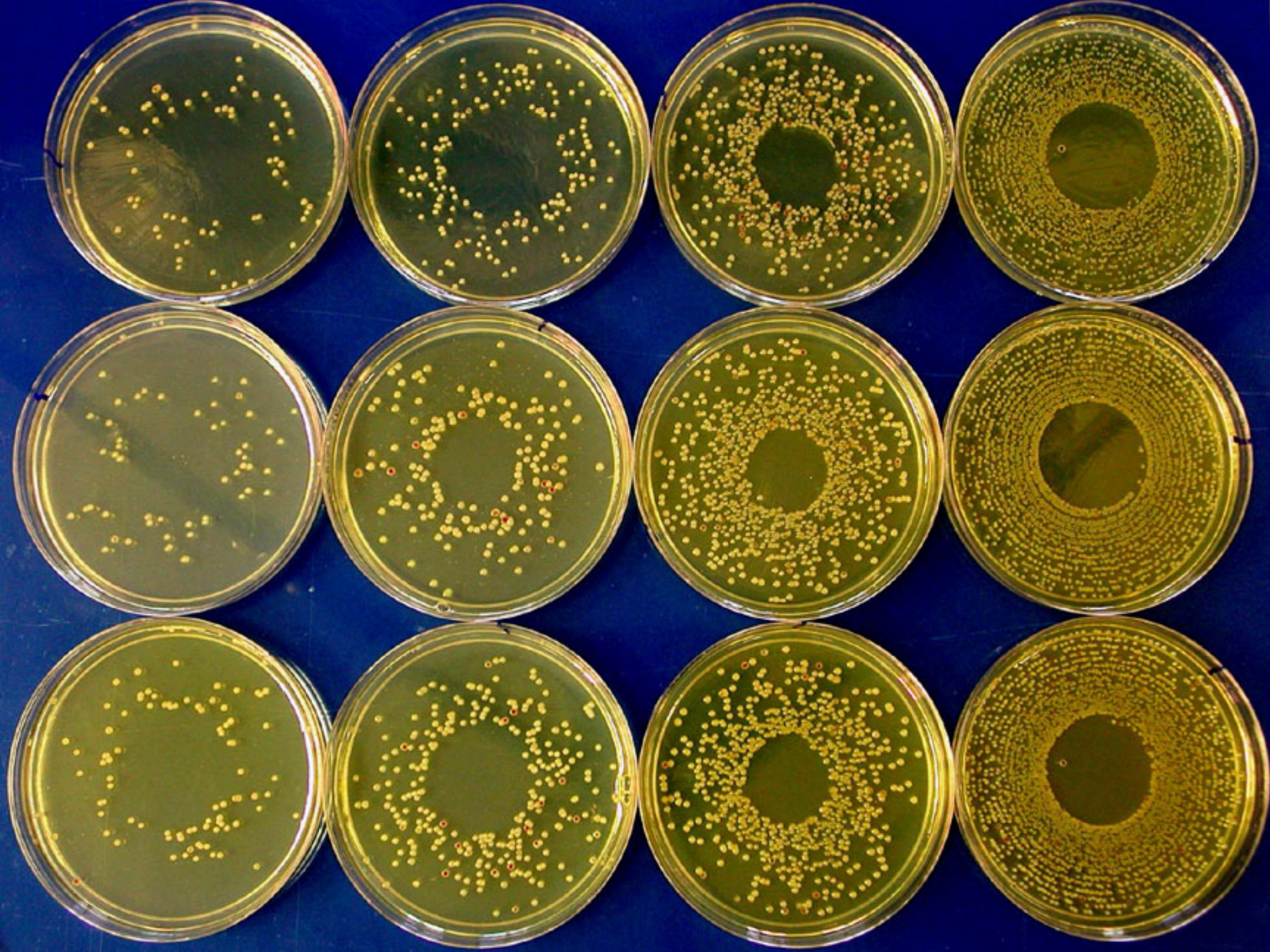
CIRCULATING HEMOCYTES



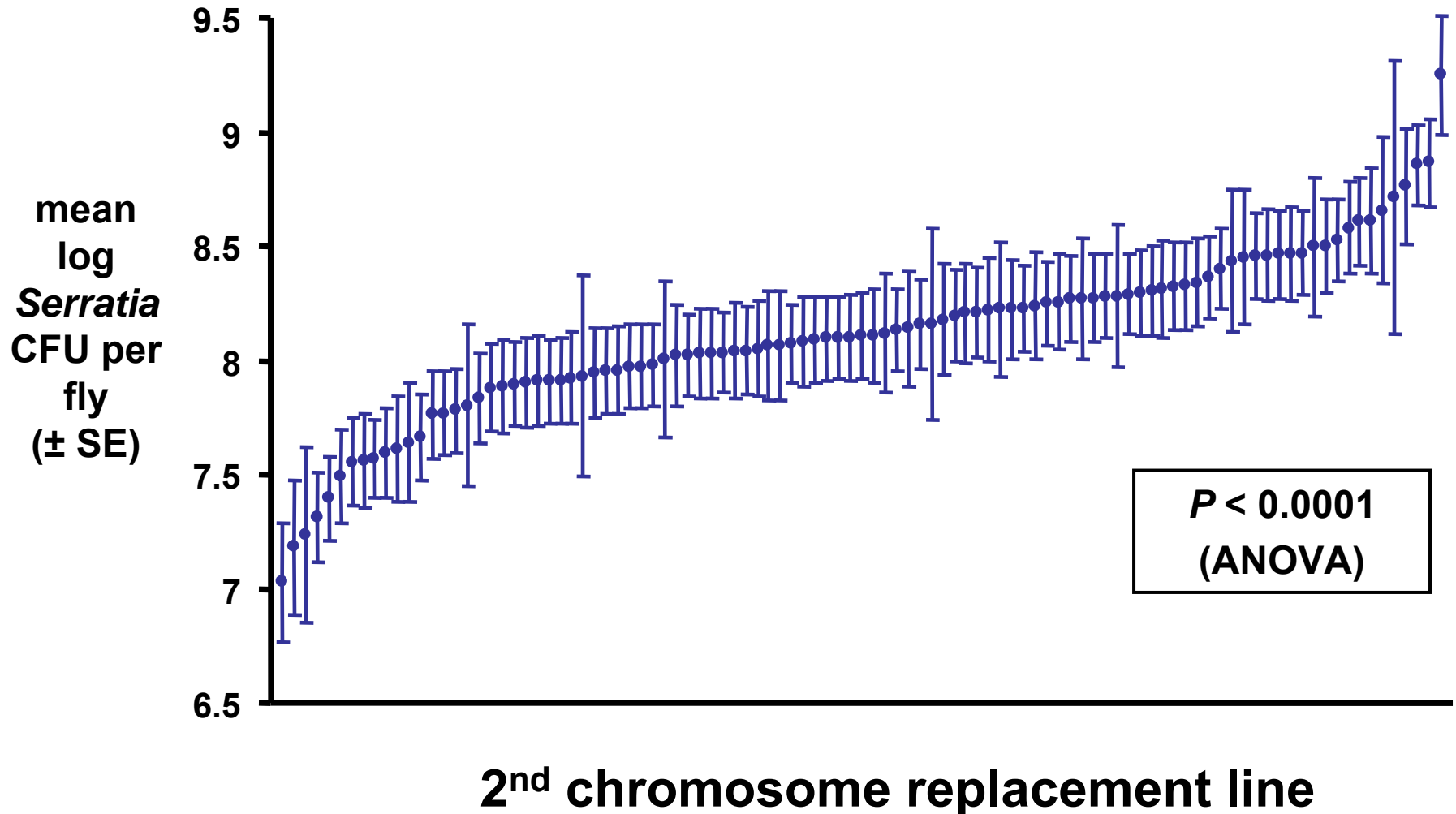
DANGER SIGNALS?



ACTIVATION of CELLULAR RESPONSE



D. melanogaster lines vary significantly in their resistance to *Serratia marcescens* infection.

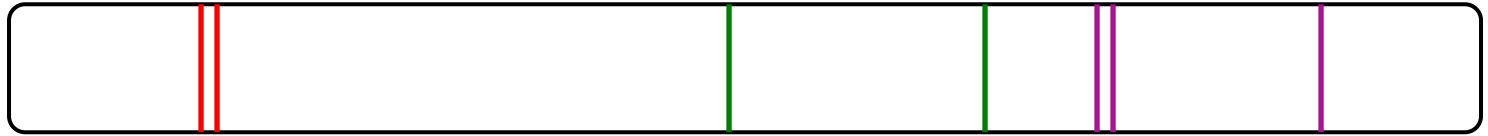


Candidate antibacterial gene loci on the *D. melanogaster* 2nd chromosome

Cytological position

2L

21 23 25 27 29 31 33 35 37 39



SC-RIV
SC-RI, III

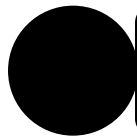
Toll-4

Tehao

cactus
Dif - dl

IK2

41 43 45 47 49 51 53 55 57 59



2R



PGRP (3)

Defensin

SC-RII

Attacin

Attacin (2)

Drosocin

Metchnikowin

Diptericin (2)

imd

Toll-7

18-Wheeler

GNBP

Pathogen recognition

Toll-like receptors

Rel signaling pathway

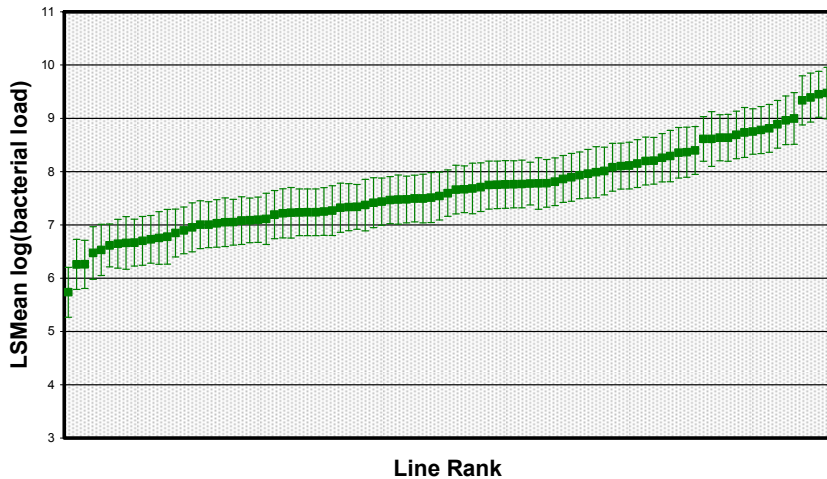
Antibacterial peptides

Most significant marker within each candidate locus

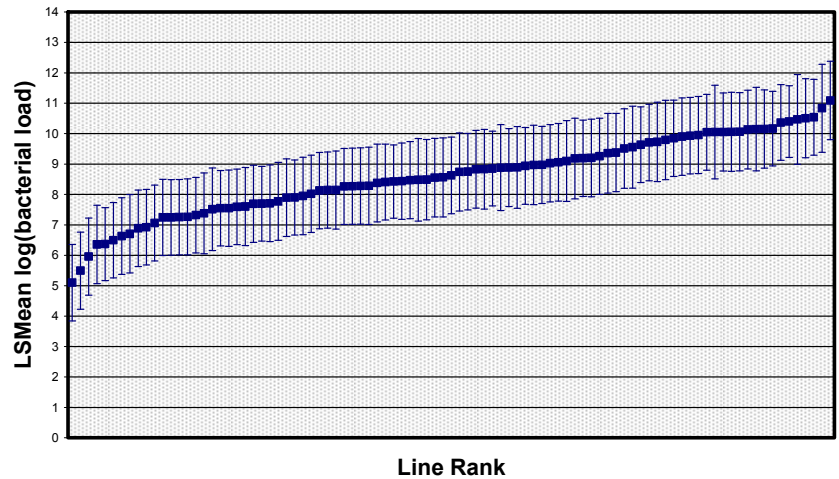
experiment-wise probability	Recognition proteins	Toll-like receptors	Signaling molecules	Antibacterial peptides
$P < 0.001$	SR-CI, PGRP-SC1B		Dif	
$0.010 < P \leq 0.001$	SR-CIV			Attacin B
$0.050 \leq P \leq 0.010$	SR-CII, PGRP-SC1A	Tehao, 18-Wheeler	cactus, imd	
N.S.	SR-CIII, PGRP-SC2	Toll-4		Attacin A, Attacin C, Diptericin A, Diptericin B, Defensin, Drosocin, Metchnikowin

Third chromosome lines also display extensive natural variation for immune function in *D. melanogaster*

Serratia marcescens



Enterococcus faecalis

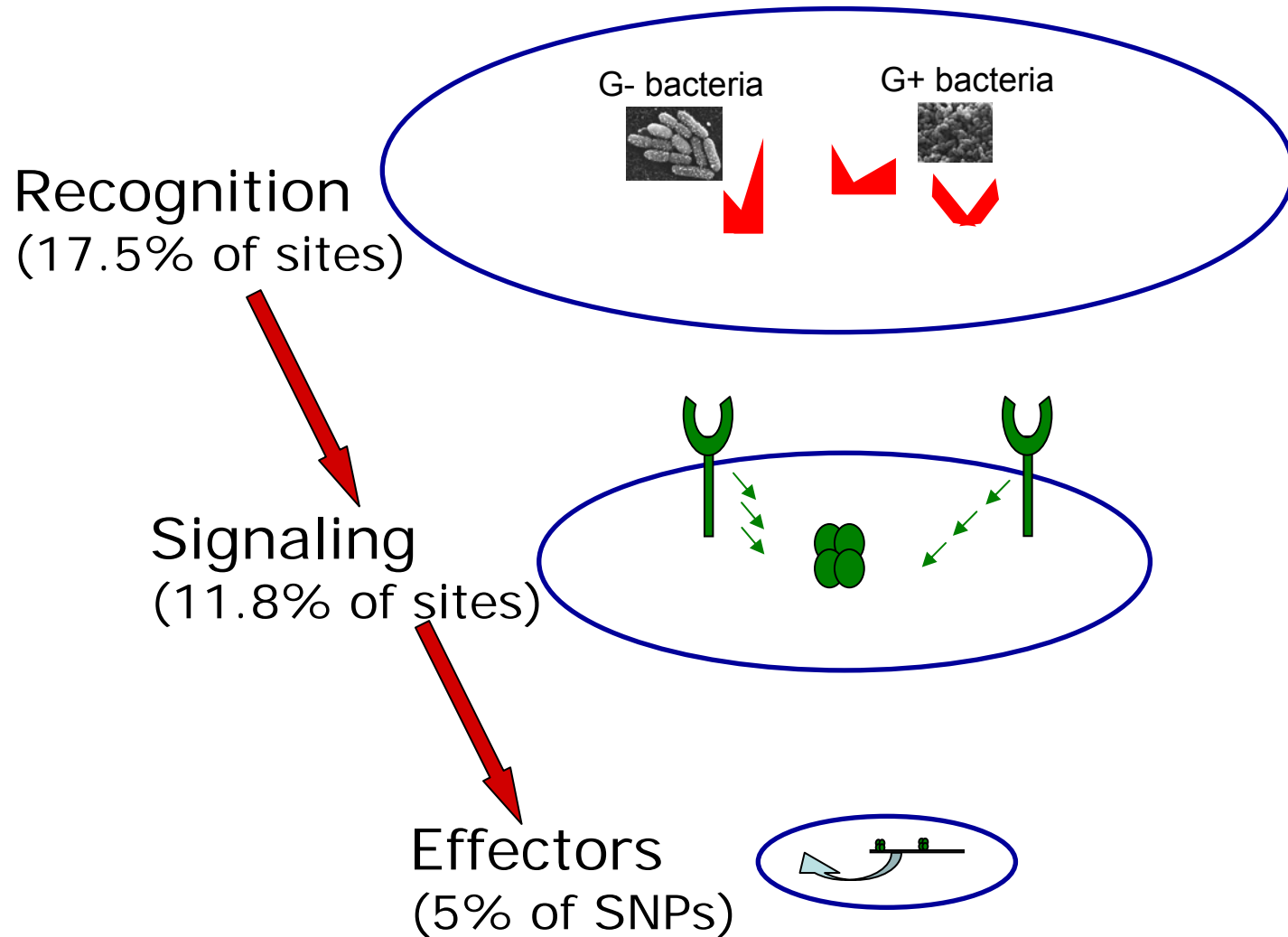


- Mean bacterial load in 3rd chromosome substitution lines 26 hours after infection with ***Serratia marcescens* (Gram negative)** or ***Enterococcus faecalis* (Gram positive)**

Associations with third chromosome recognition proteins

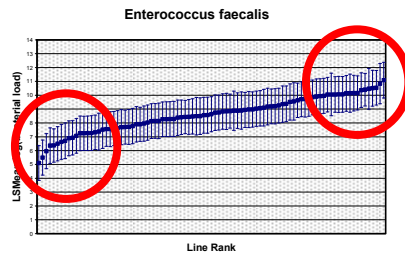
		Gram+		Gram-	
		Ef	LI	Pr	Sm
GNBP1	intron	0.7586	0.3420	0.0131	0.1544
GNBP1	intron	0.8942	0.0351	0.0076	0.0525
GNBP1	5' UTR	0.9727	0.0491	0.8596	0.5135
GNBP1	5' UTR	0.9600	0.4002	0.0179	0.1965
GNBP1/2	intergenic	0.7870	0.2431	0.0201	0.3019
GNBP2	exon	0.7974	0.0480	0.1570	0.1217
GNBP3	upstream	0.2197	0.0394	0.0106	0.0224
GNBP3	upstream	0.4018	0.1250	0.0066	0.0063
PGRP-LA	exon	0.1007	0.5285	0.0038	0.4228
PGRP-LA	exon	0.9966	0.0330	0.8913	0.9324
PGRP-LA	exon	0.6958	0.9759	0.2744	0.0493
PGRP-LB	exon	0.0780	0.0466	0.2278	0.6012
PGRP-LC	upstream	0.0372	0.9185	0.2237	0.6046
PGRP-LC	upstream	0.0089	0.2625	0.8778	0.3352
PGRP-LD	5' UTR	0.3784	0.7141	0.2042	0.0026
PGRP-LD	intron	0.8134	0.2948	0.8090	0.0265
PGRP-LF	intron	0.0285	0.1461	0.5360	0.9988
PGRP-LF	3' UTR	0.3134	0.0051	0.7673	0.0499
PGRP-LF	upstream	0.7276	0.0077	0.0661	0.3127
PGRP-SB1	upstream	0.1541	0.0633	0.0055	0.9493
PGRP-SB2	intron	0.0033	0.2062	0.8989	0.1660
PGRP-SB2	exon	0.0136	0.2470	0.7951	0.2760
PGRP-SD	exon	0.2805	0.9817	0.0013	0.3318
PGRP-SD	upstream	0.2331	0.6324	0.0143	0.7719

SNPs: significant associations with immune competence



Measuring gene expression: experimental design

Pick lines at the phenotypic extremes
(n=30)

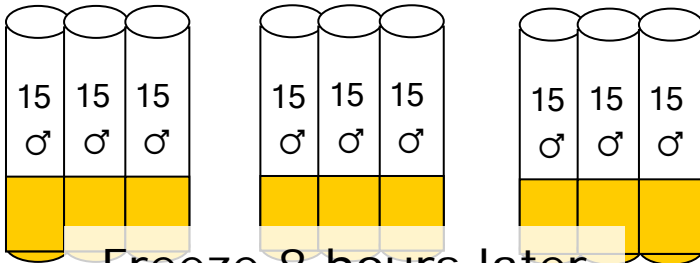


For each line:

Serratia
infected

Enterococcus
infected

Naïve



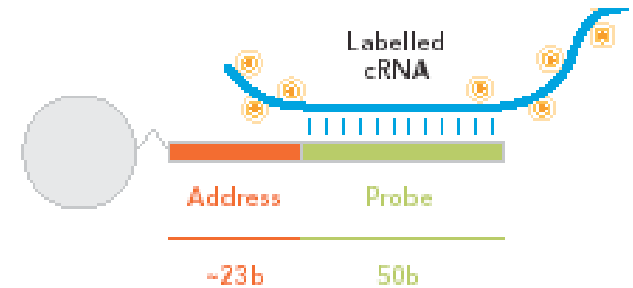
For each sample (n=270):

Extract RNA

Synthesize cDNA

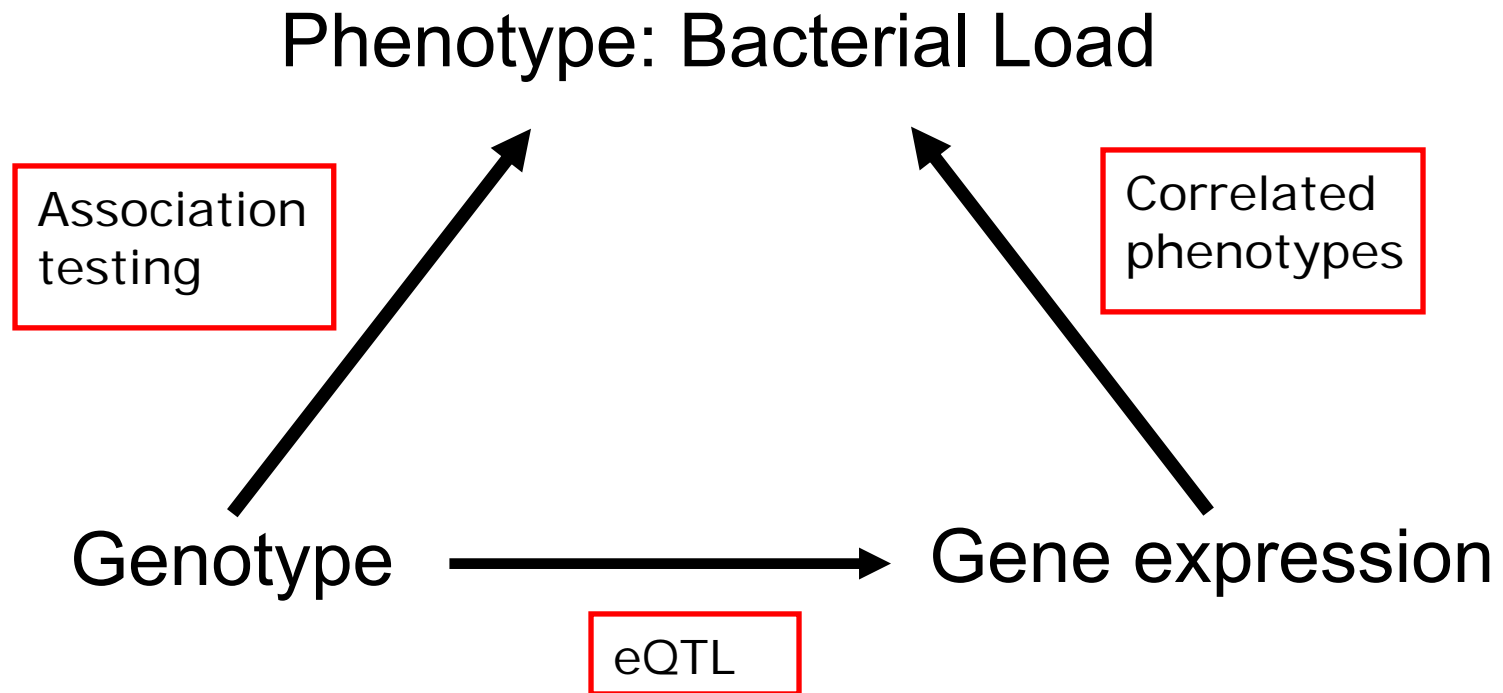
Synthesize and label cRNA

Hybridize to Illumina Beadchip

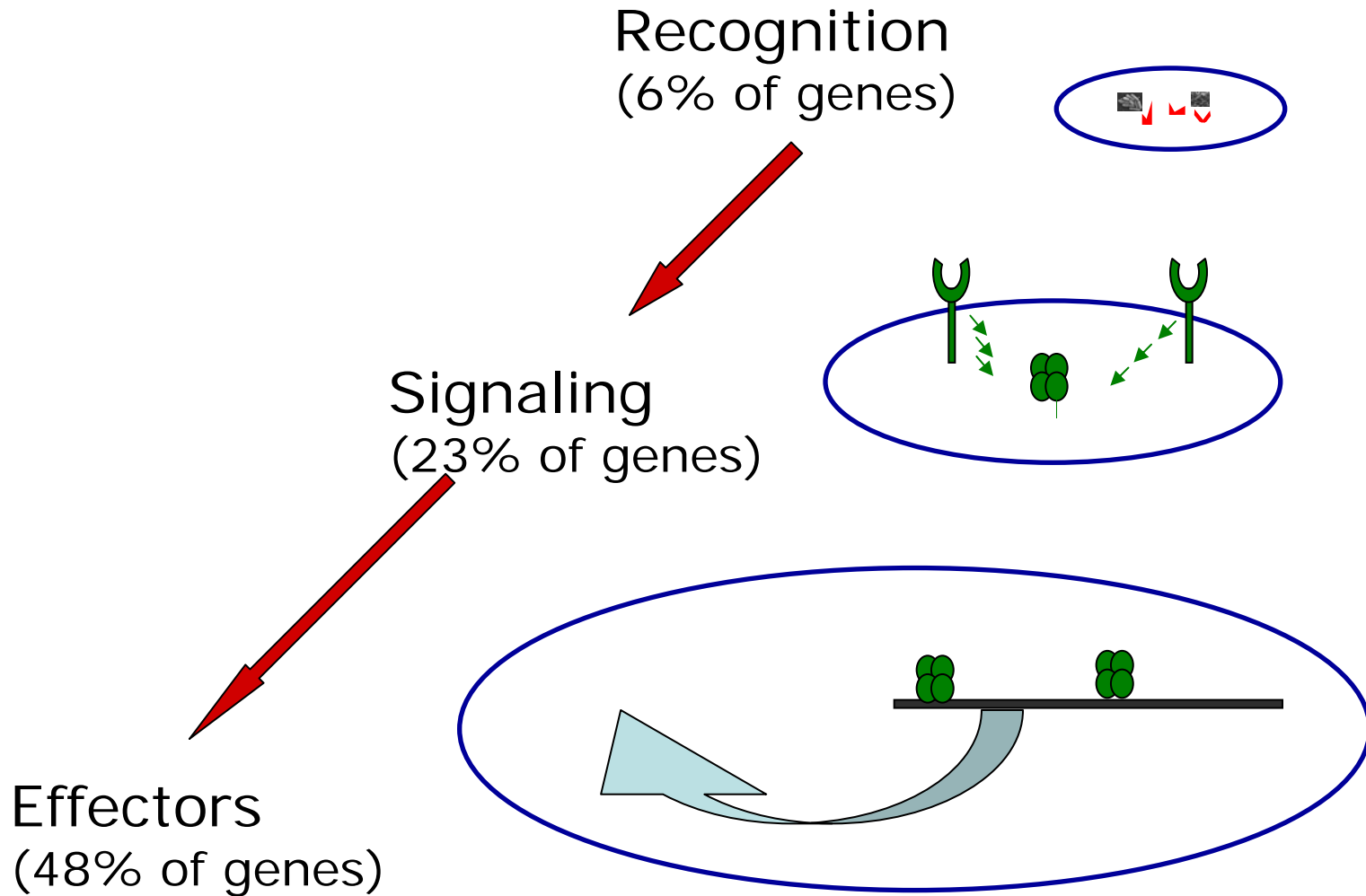


Log transform expression data
Normalize using quantile spline

Predicting Immune competence



Gene expression: significant associations with immune competence



Genotype-phenotype conclusions

- Every chromosome harbors significant variation in immune competence.
- Immune competence across different pathogens is largely uncorrelated.
- SNPs in recognition and signaling proteins are often associated with bacterial load.
- Variation in expression of antimicrobial peptides significantly predict bacterial load.

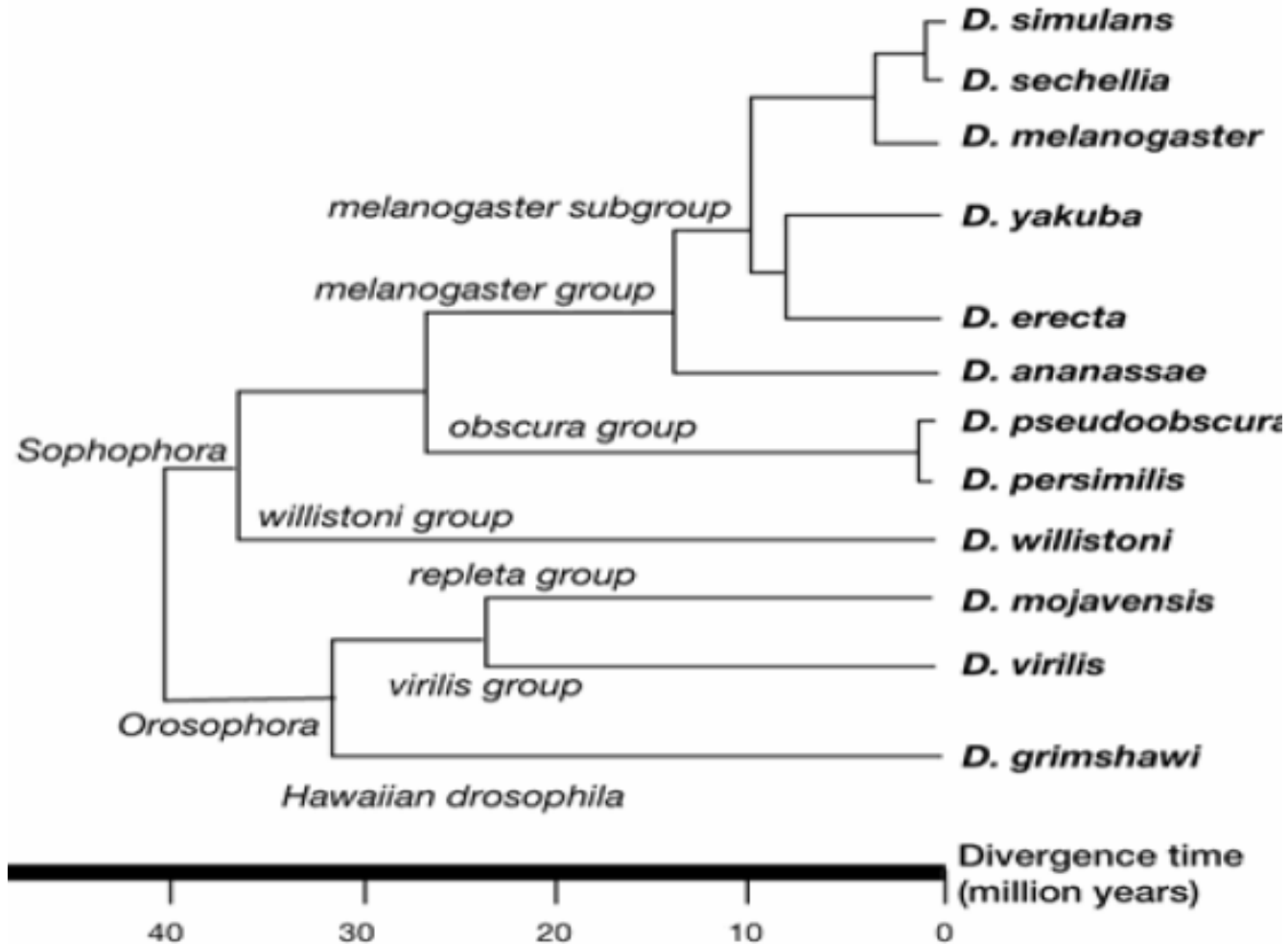
In the works...

- X chromosome replacement lines
- Multiple pathogens, including wild-derived
- Type III secretion mutant *Pseudomonas*
- microRNA induction and regulation
- Deficiencies and quantitative complementation
- RNAi knockdowns of key transcripts
- Diallel crosses of extreme lines
- F1 hybrids and interspecific introgressions

Comparative genomics and evolution of immunity

- Genome sequences of
 - Diptera (Drosophila)
 - Hymenoptera (honeybee)
 - Coleoptera (flour beetle)
 - Lepidoptera (silk moth)
- All have Toll, imd, JNK, and JAK/STAT pathways.
- Large differences in gene families.
- But too far diverged to say much more.

Completed *Drosophila* genome sequences (AAAwiki)



Molecular evolution as a tool for understanding innate immunity

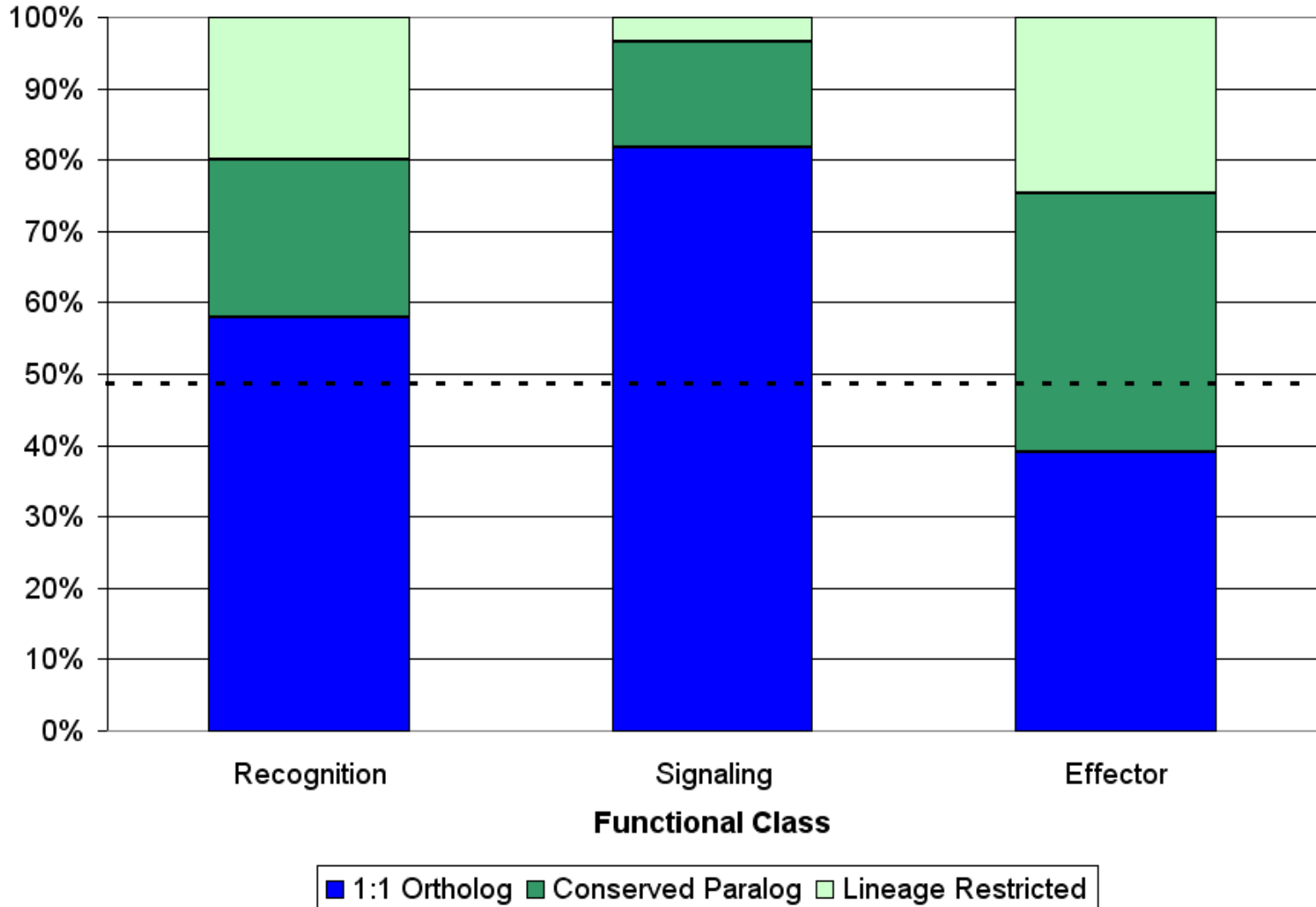
- Immune systems frequently show coevolutionary arms races.
- We can fit molecular evolutionary models to identify rapidly evolving genes.
- Where is natural selection impacting the long term evolution of the pathway?

What is the D matrix?

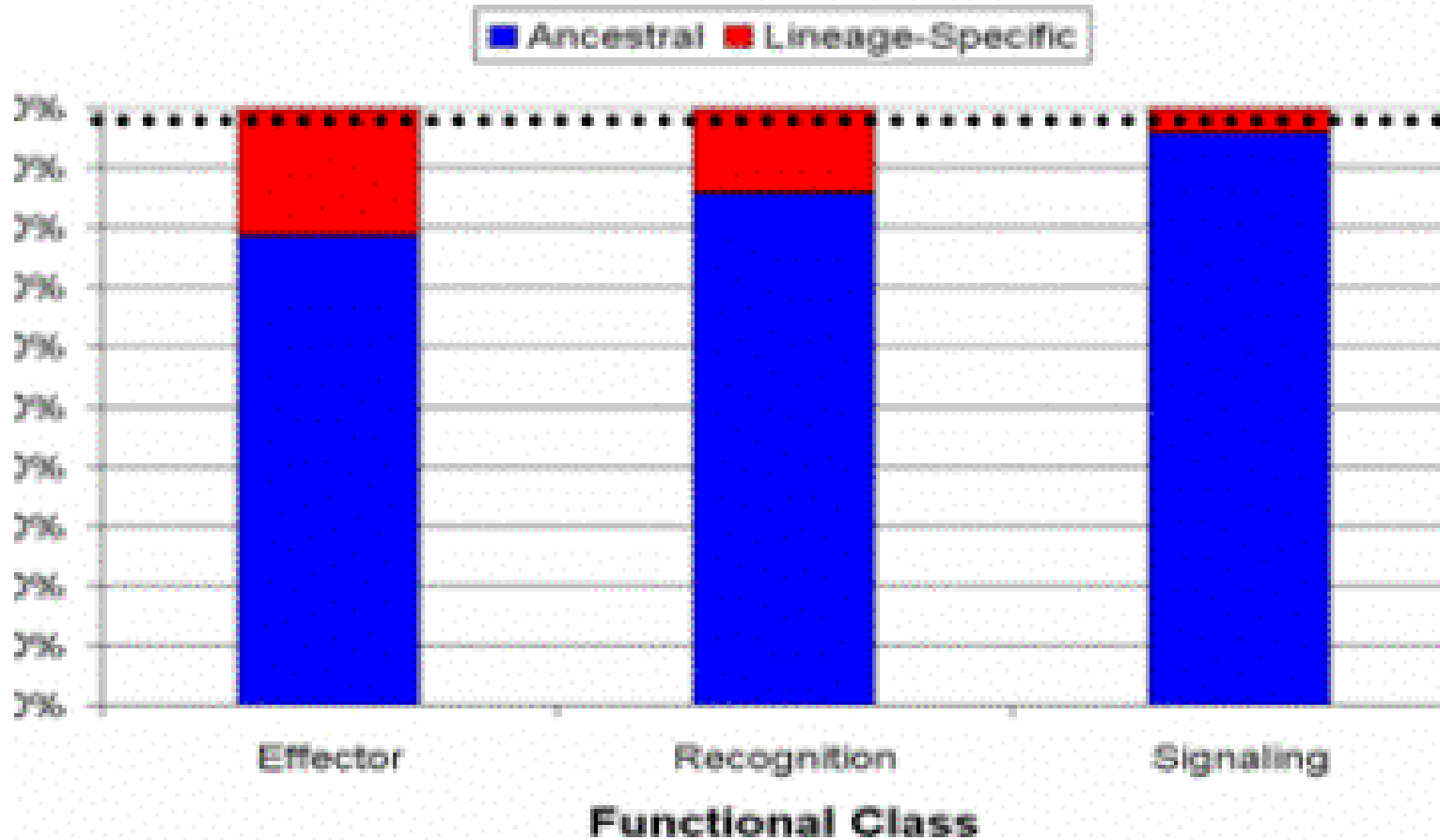
Method for studying comparative genomics of innate immunity

- Identify 248 genes in *D. melanogaster* involved in innate immunity.
- Pull genes from paralogy networks constructed by fuzzy reciprocal BLAST of GLEANR models.
- Re-align with TCOFFEE.
- Mask problem regions.
- Annotate 2474 gene models that are orthologs and paralogs of the melanogaster genes.
- Estimate divergence rates, fit codon substitution models, fit birth-death models, etc..

Signalling proteins are more likely to occur as one-to-one orthologs across the phylogeny

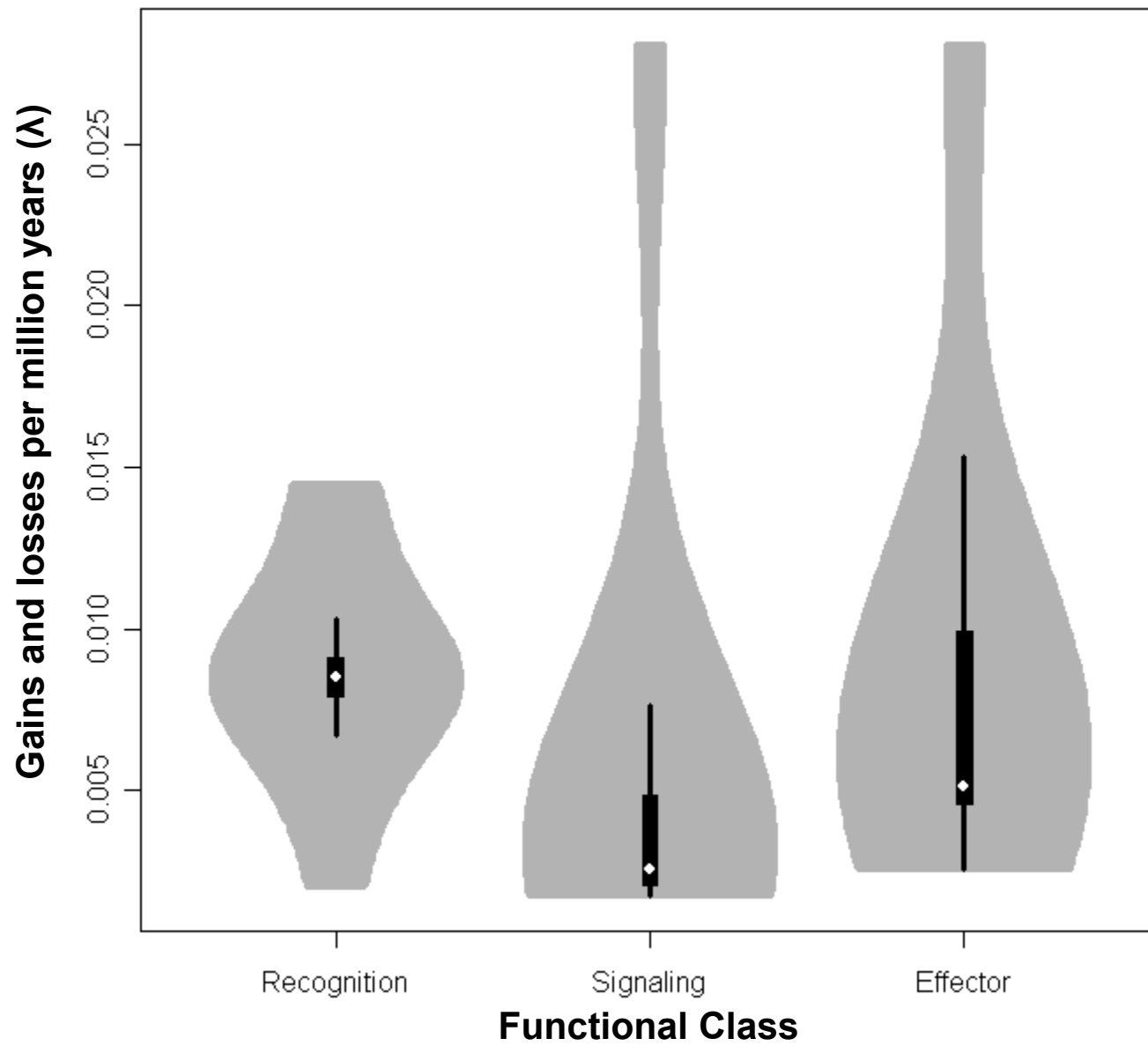


Effector proteins have the highest proportion of lineage-restricted genes

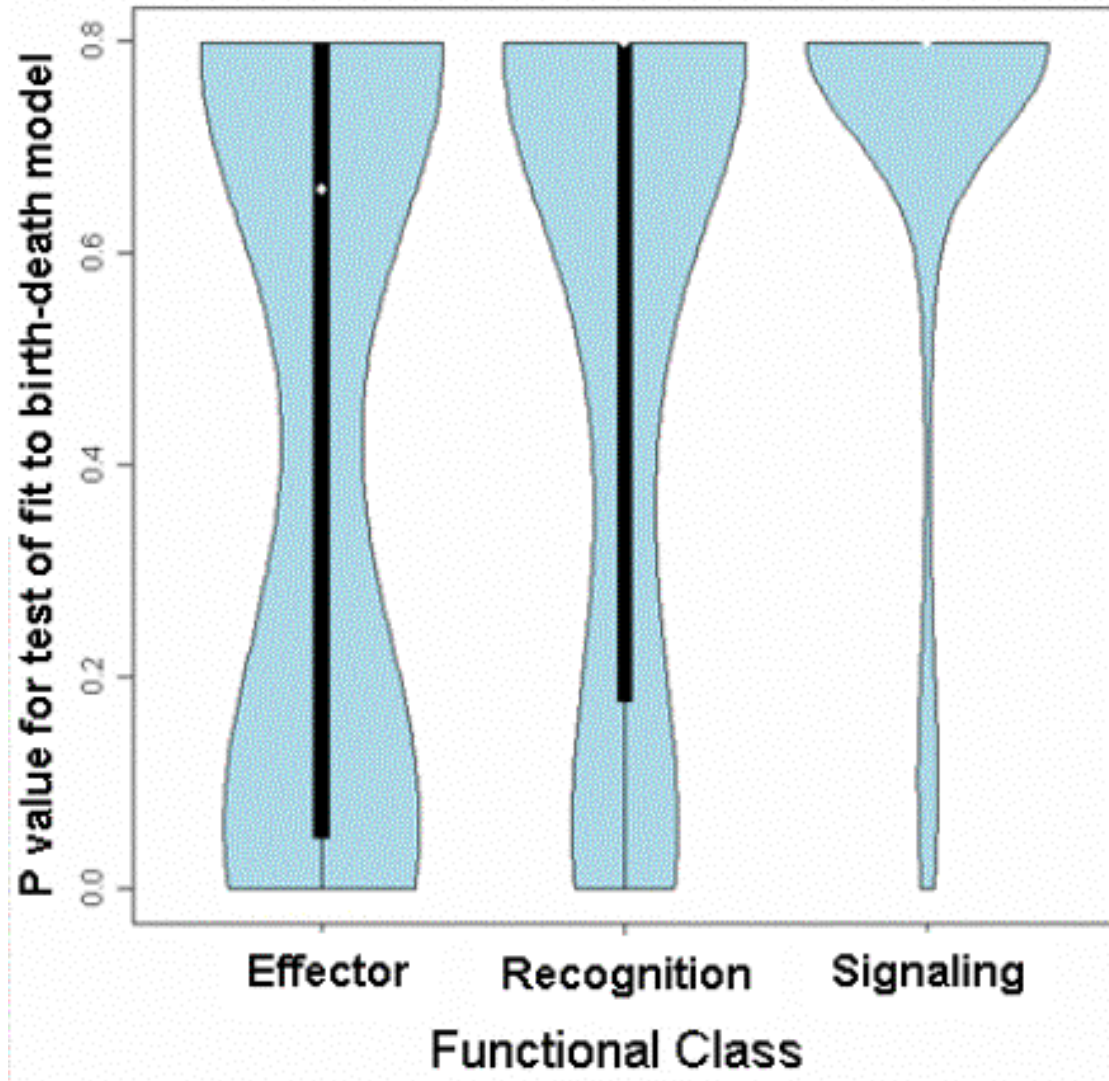


Test homogeneity of birth-and-death

- Across a phylogeny, observe different counts of genes within a gene family.
- Are the counts consistent with a homogeneous birth-death process?
- CAFÉ fits to an equilibrium model with rate λ .
- Assess significance by likelihood ratio to alternative model with different rates.



Many effector proteins reject the birth-death model
Signalling proteins do not.



Classes of recognition peptides

- Thioester-containing proteins that act like complement factors (TEPs)
- Membrane-bound hemocyte-specific receptors (*e.g.*, eater, Nimrods, scavenger receptors with a role in phagocytosis)
- Secreted or membrane-bound recognition proteins that recognize conserved bacterial cell wall components (*e.g.*, PGRPs and GNBPs).

Conserved role of PGRPs

- PGRPs are conserved as one-to-one orthologs across the entire *Drosophila* phylogeny, suggesting that the expansion of the PGRP family predates the deepest branch in this phylogeny.
- Among the PGRPs, the only genes that have a history of duplication in *Drosophila* are the PGRP-SC and PGRP-SB families

Major signalling pathways

- Toll
- Imd
- JAK/STAT
- JNK
- p38 stress response pathway
- Hemocyte development, differentiation, and proliferation
- Nuclear pore proteins involved in Nf- κ B translocation
- Other miscellaneous signaling proteins

Nearly perfect one-to-one orthology of signalling peptides

- An exception: DIF, an Nf- κ B transcription factor distantly related to dorsal, is absent from mosquitoes and honeybees.

Toll-like receptors

- Toll, Toll-6, Toll-7, Toll-9, Tehao, Tollo and 18w all exist as single orthologs across all 12 species
- But Toll-4 shows an abundance of lineage-specific changes in number.

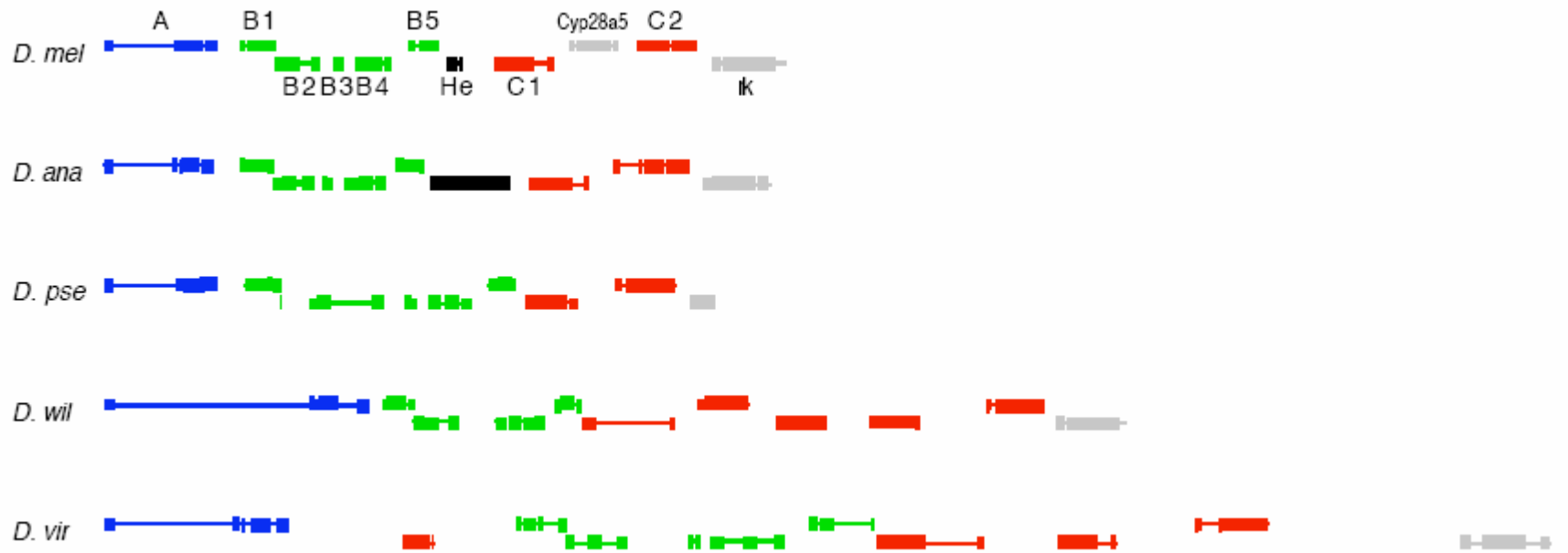
Humans have 10 Toll-like receptors

Sea urchin has more than 300 !!

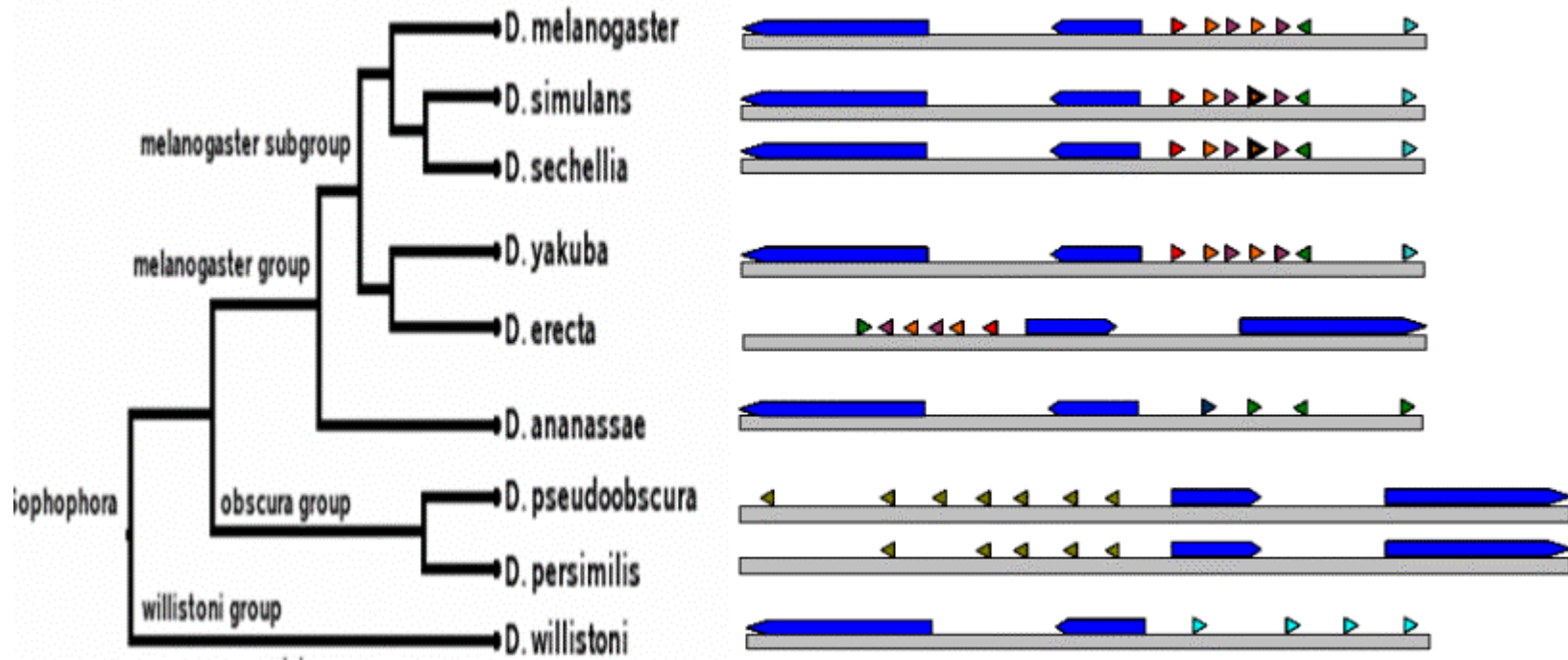
Classes of effector peptides

- Antimicrobial peptides
 - Cecropins***
 - Drosomycins***
 - IM1 cluster***
- Transferrins
- Turandot proteins***
- Lysozymes***
- Phenoloxidase cascade
- Proteins involved in coagulation
- Gut immunity

*** reject CAFÉ model of homogeneous birth-and-death



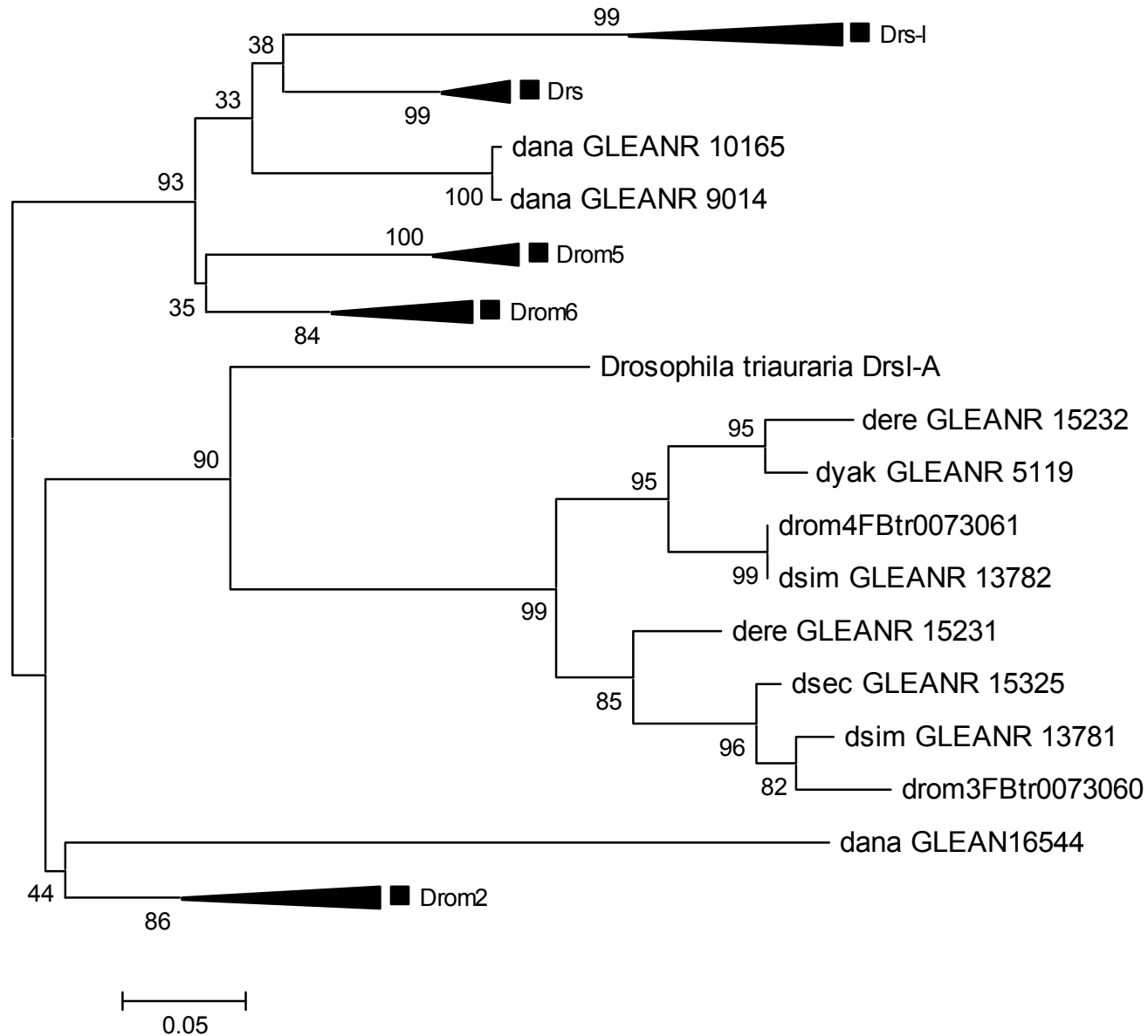
The cecropin gene family



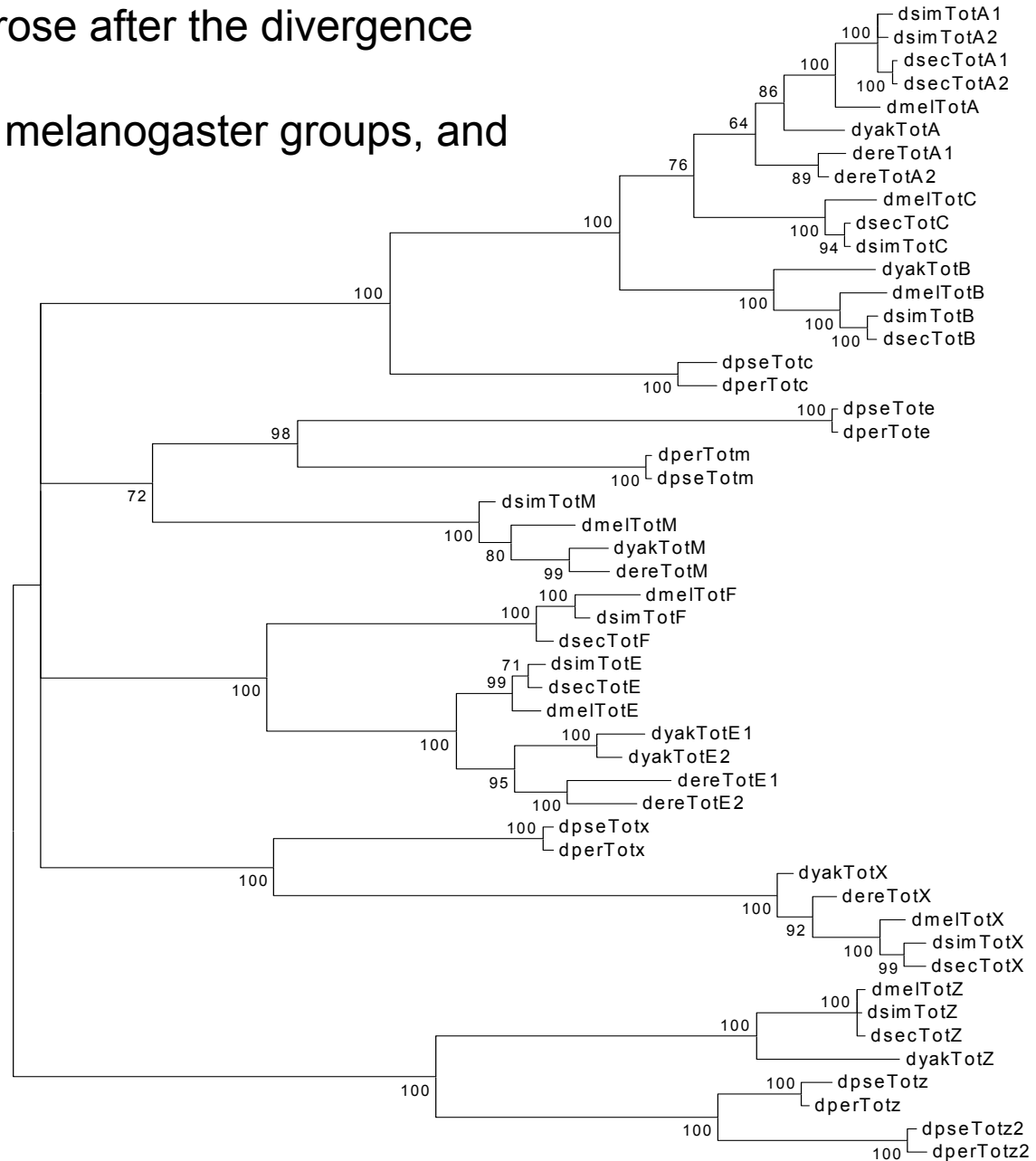
Turandots and the Drosomyctins

- Present in only a subset of the lineages on the phylogeny.
- Appear to have duplicated rapidly shortly after arising.
- Birth-death model appears to be rejected because of the rapid expansion of the families soon after they appeared

Drosomyctins appear only in the melanogaster group



The Turandot family arose after the divergence of the willistoni group from the obscura and melanogaster groups, and was subsequently lost in *D. ananassae*

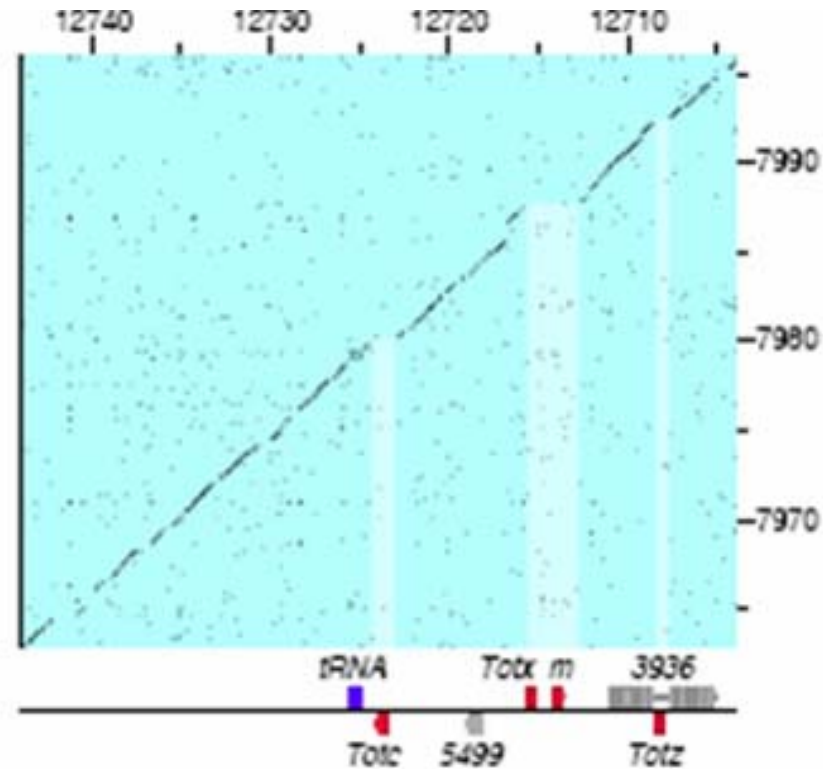


Expected changes per site

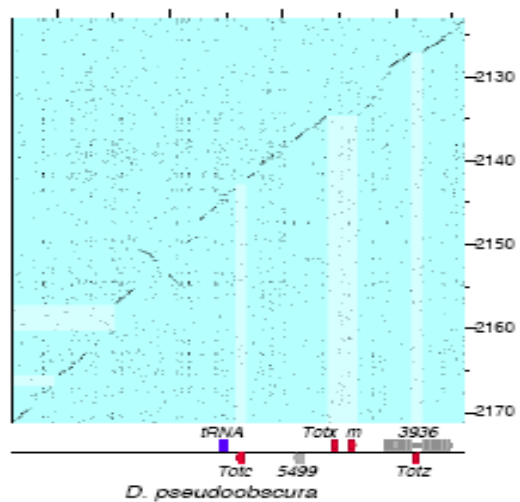
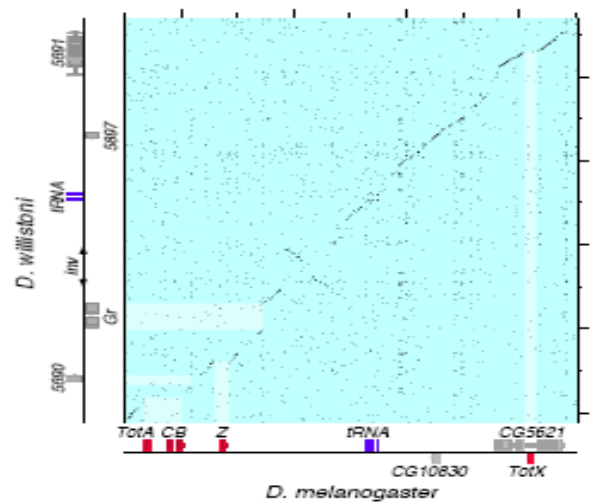
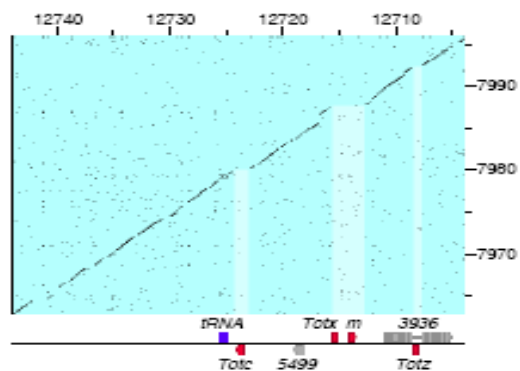
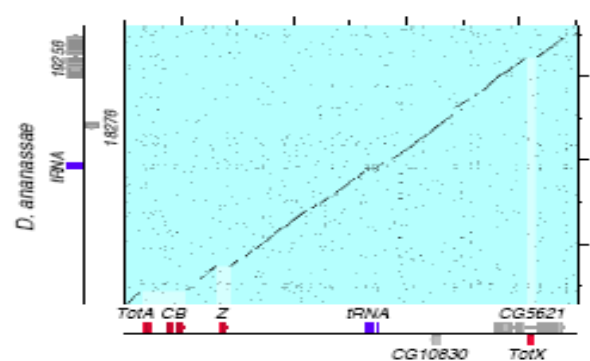
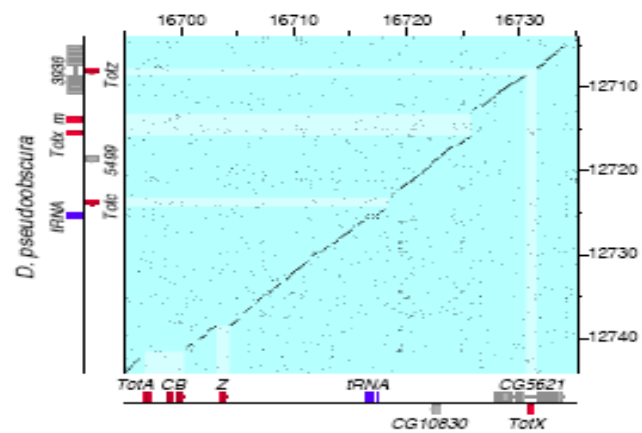
0.5

Turandot gain and loss

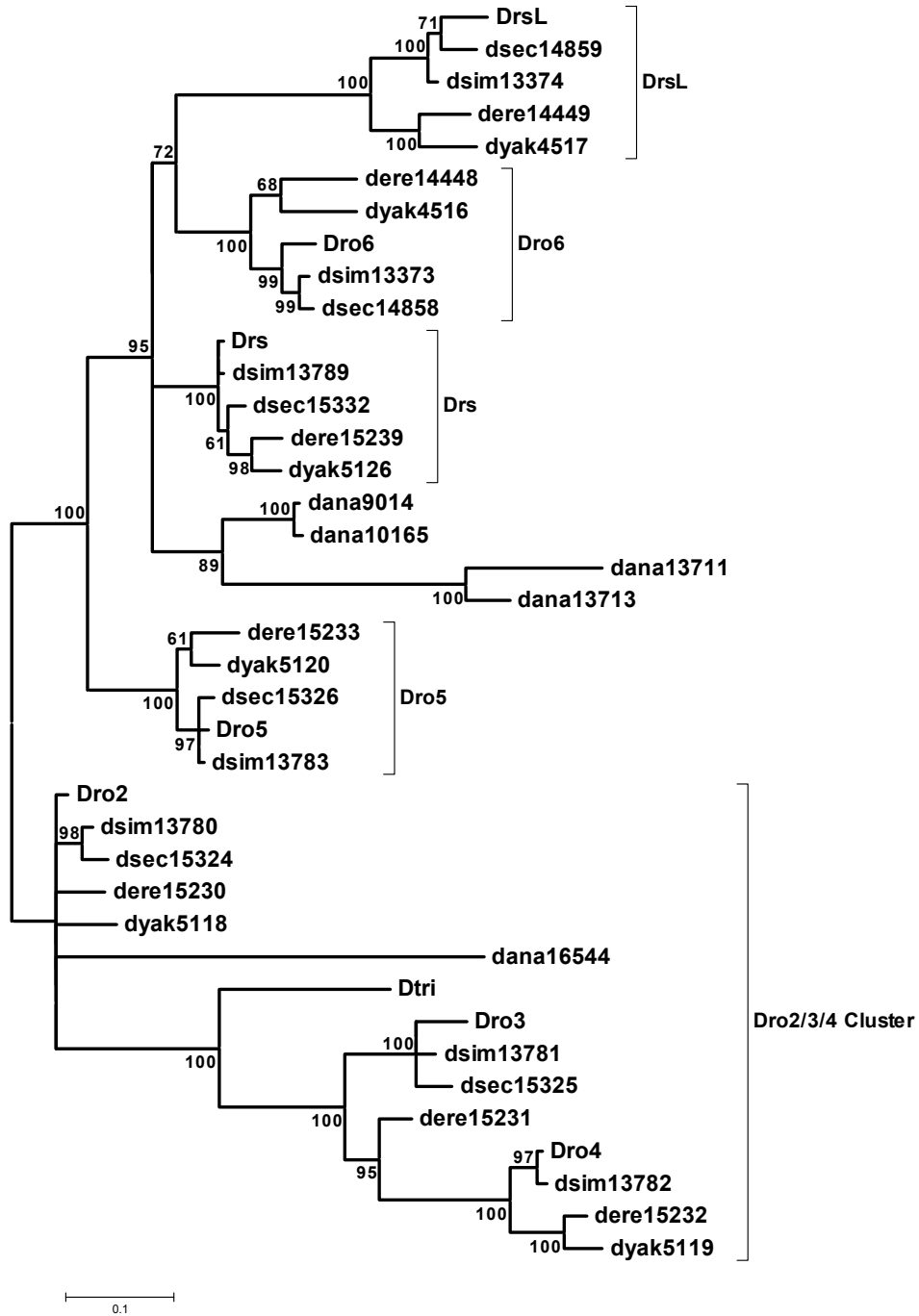
D. ananassae



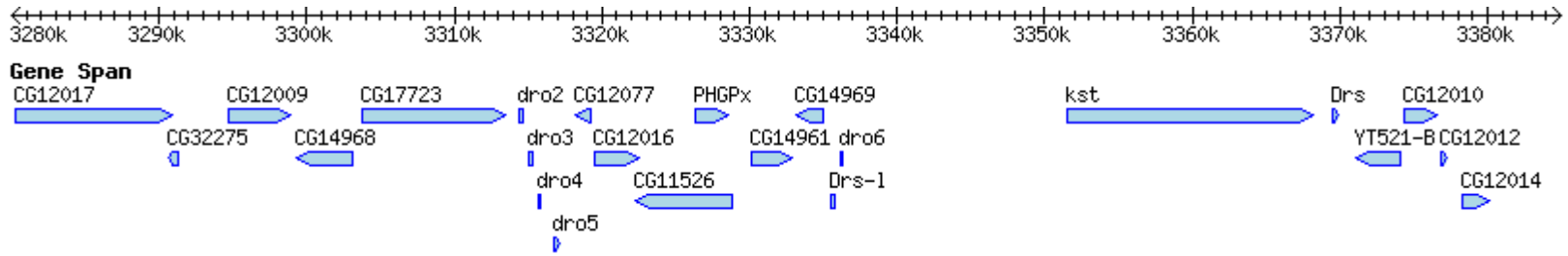
D. pseudoobscura



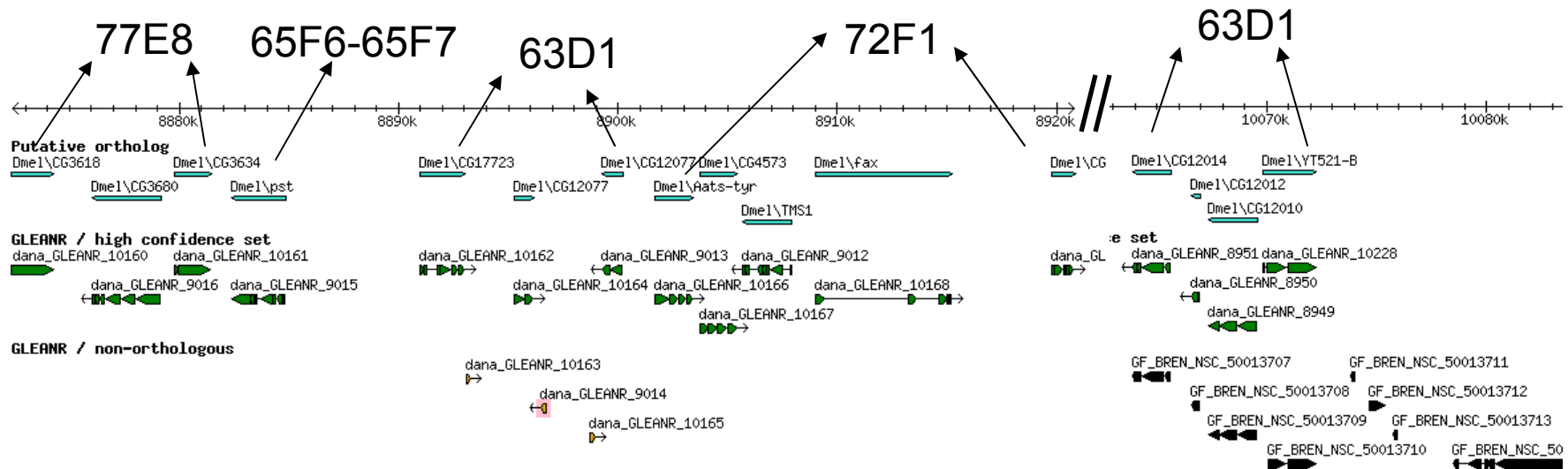
A

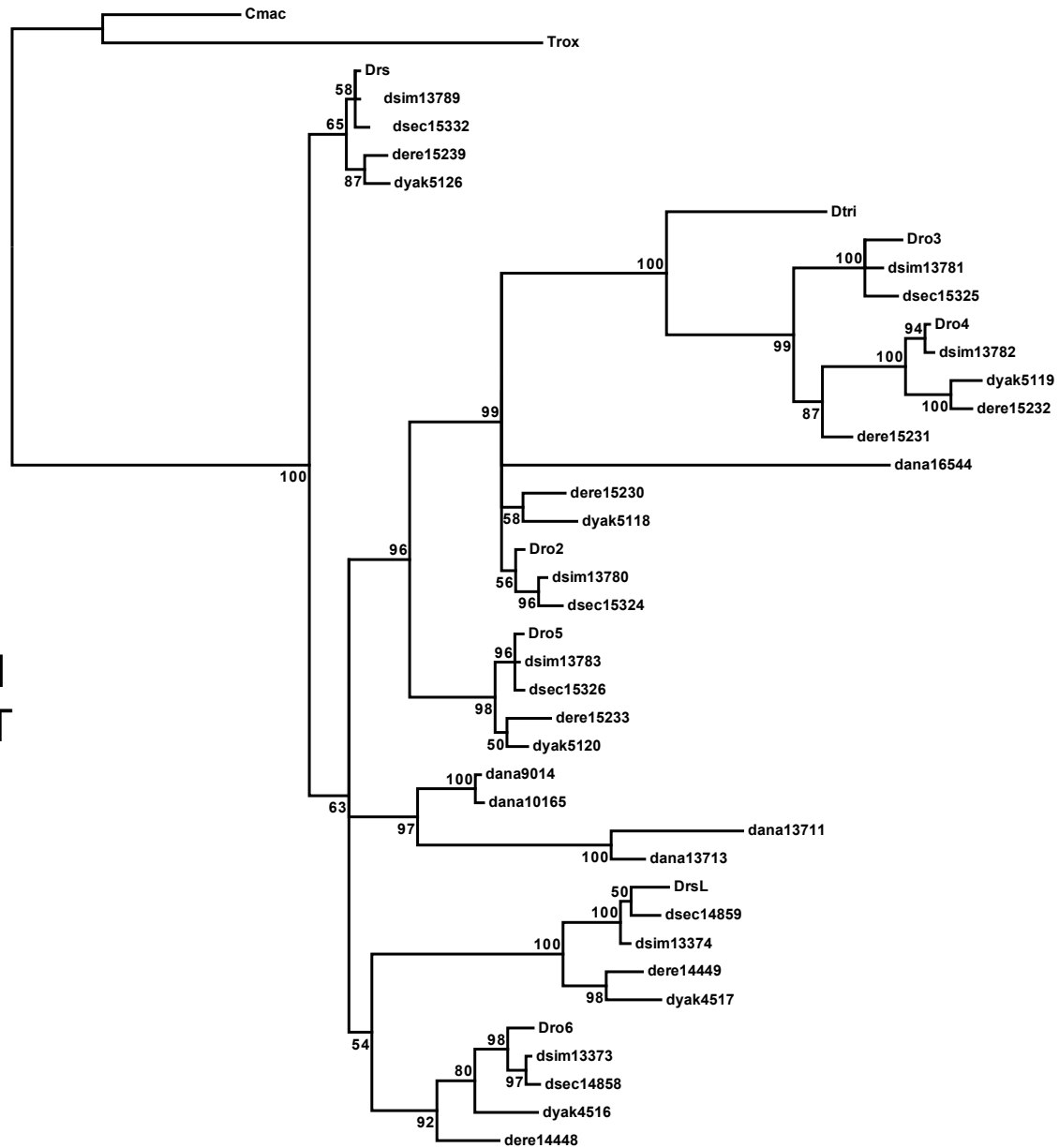


Genomic organization of *Drs* cluster in *D. melanogaster*



Genomic organization of *Drs* cluster in *D. anannasae*, along with the location in *D. melanogaster* that each gene maps.

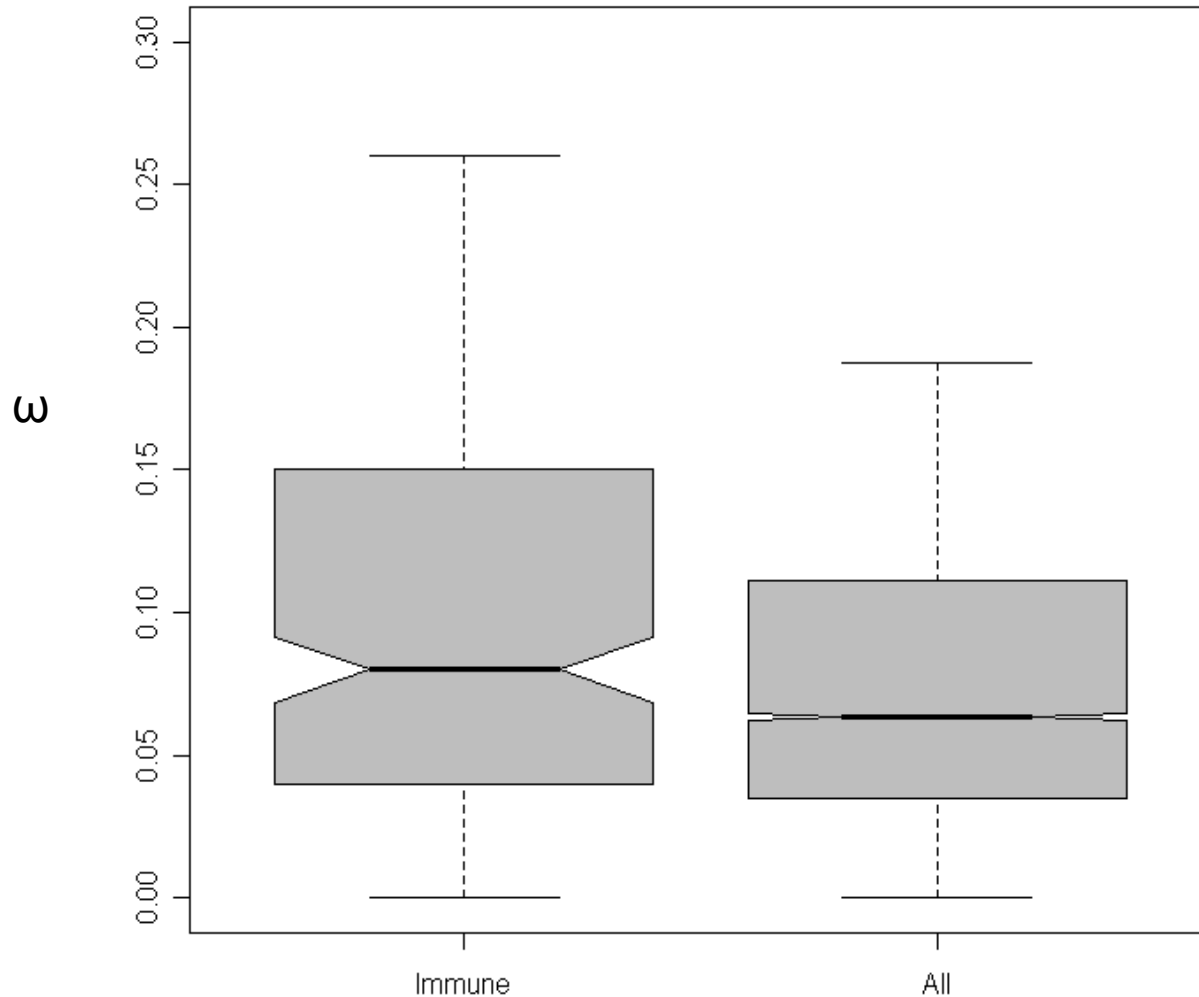




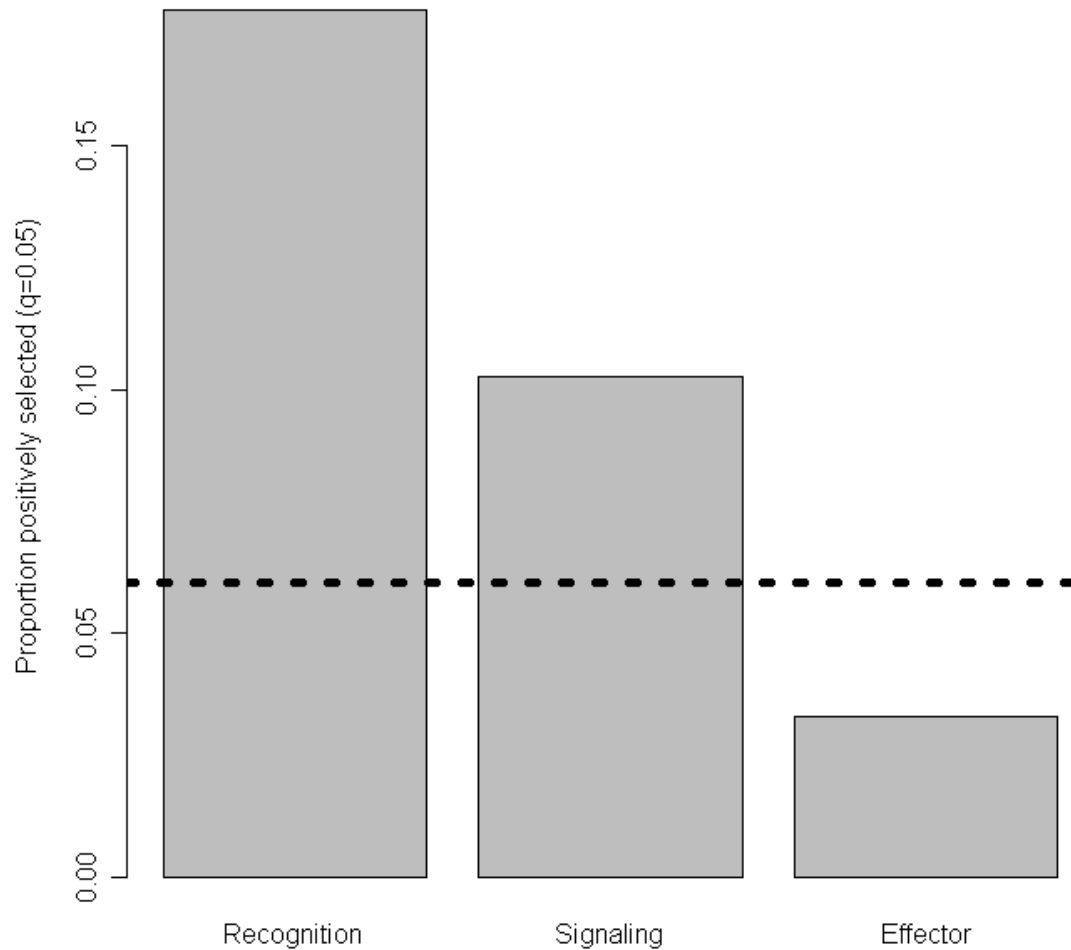
Including, and rooted
with coleopteran EST
sequences

0.1

Immune proteins are significantly less conserved than the proteome as a whole



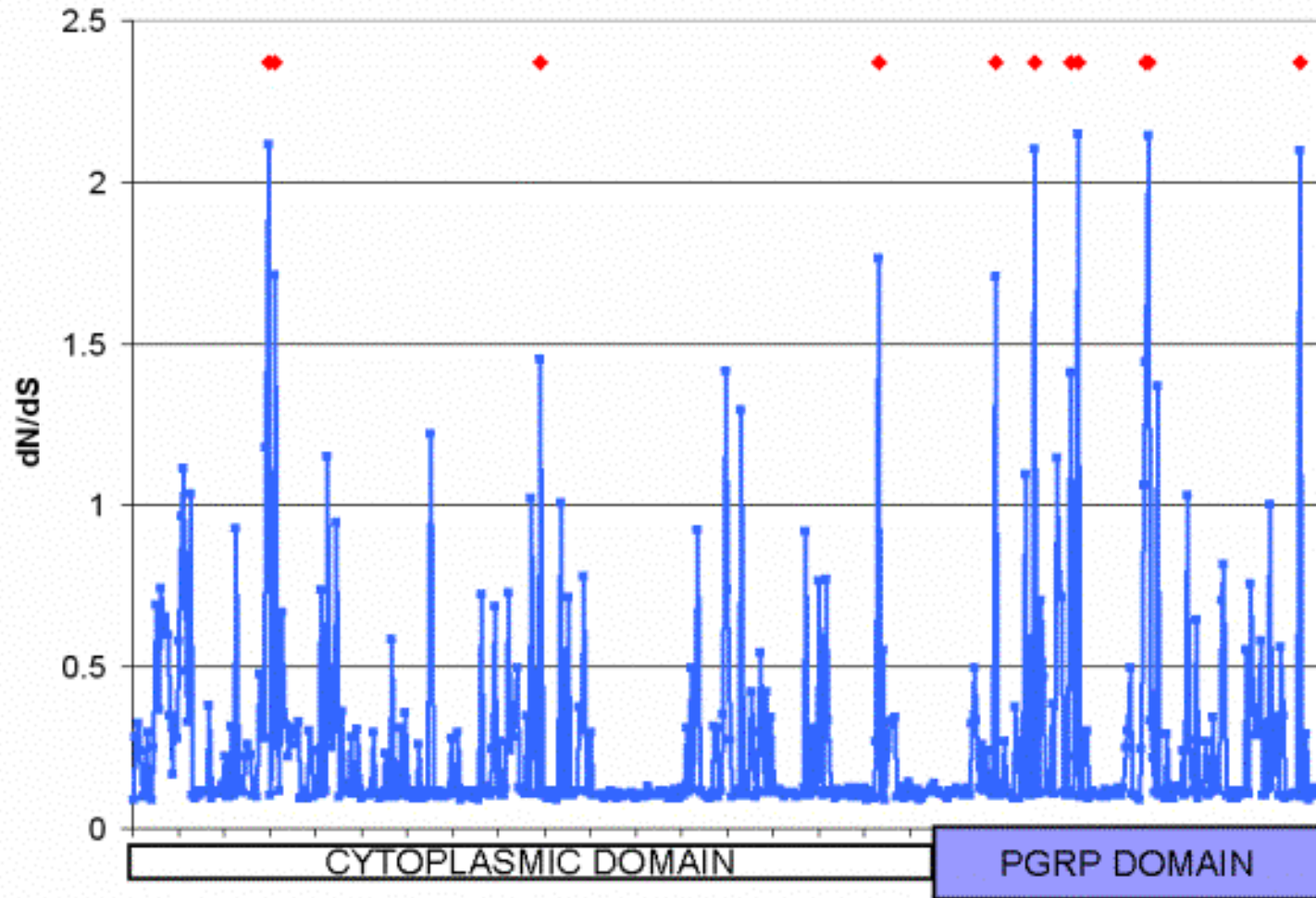
Significant differences in the proportion of positively selected genes



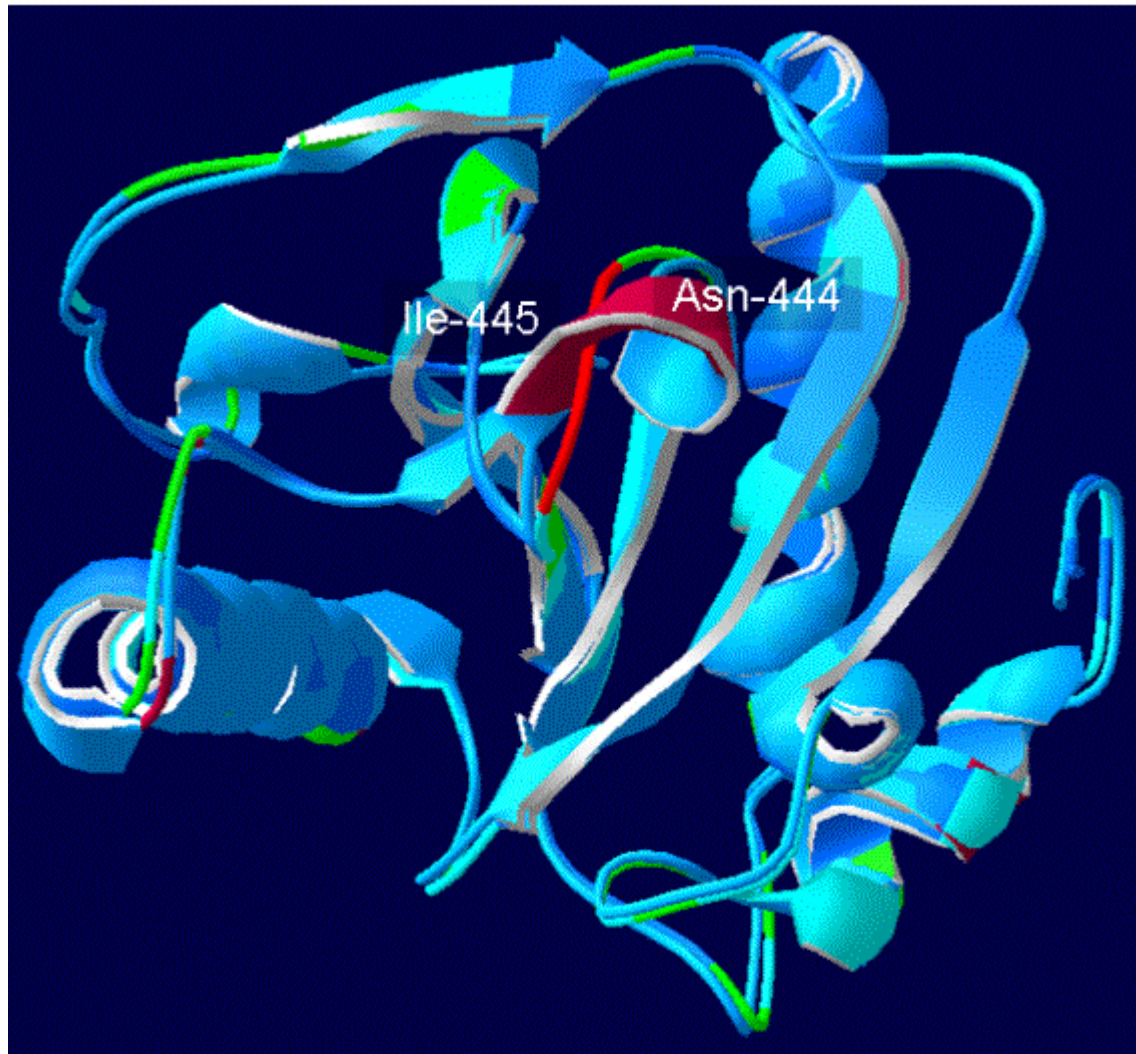
Positive selection in PGRP-LC

- PGRP-LC is alternatively spliced in *Drosophila melanogaster* to produce three isoforms with different PGRP domains.
- All three splice forms, designated PGRP-LCa, -LCx, and -LCy are conserved in all twelve species studied, so we analyzed each isoform separately.
- Only PGRP-Lca shows evidence for positive selection.

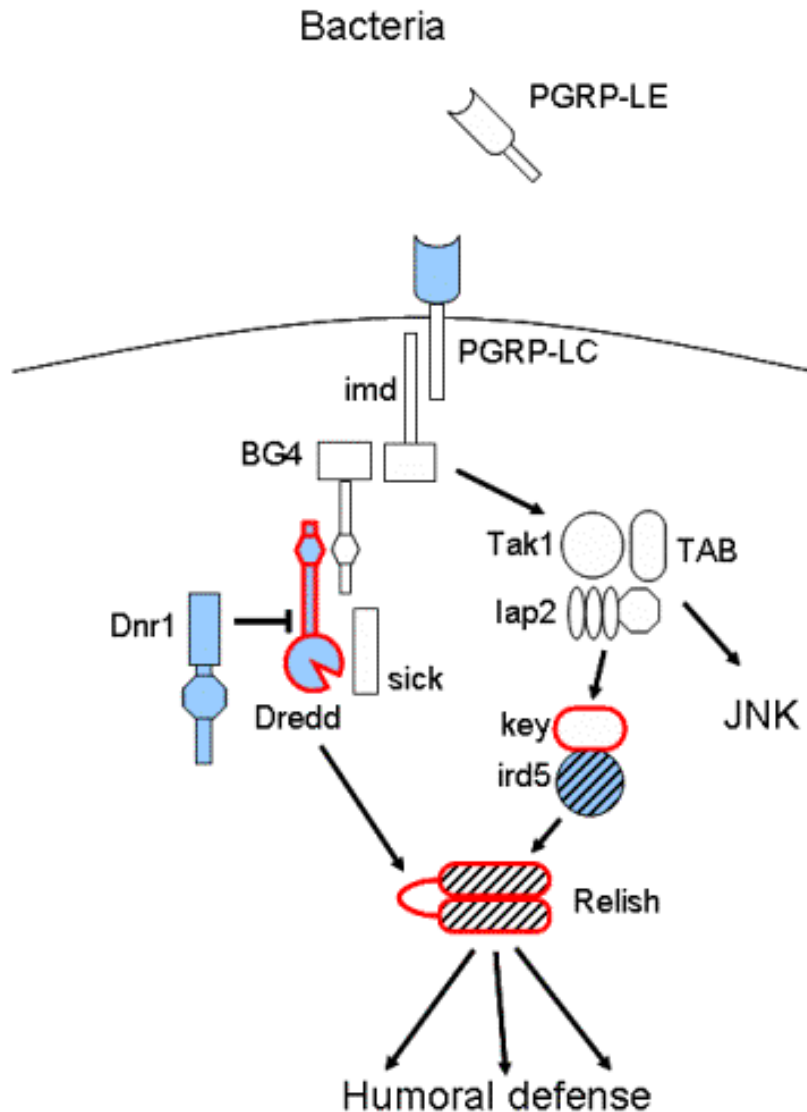
PGRP domain is target of positive selection in PGRP-LCa



The melanogaster subgroup only has a 2 amino acid insertion in PGRP-LC



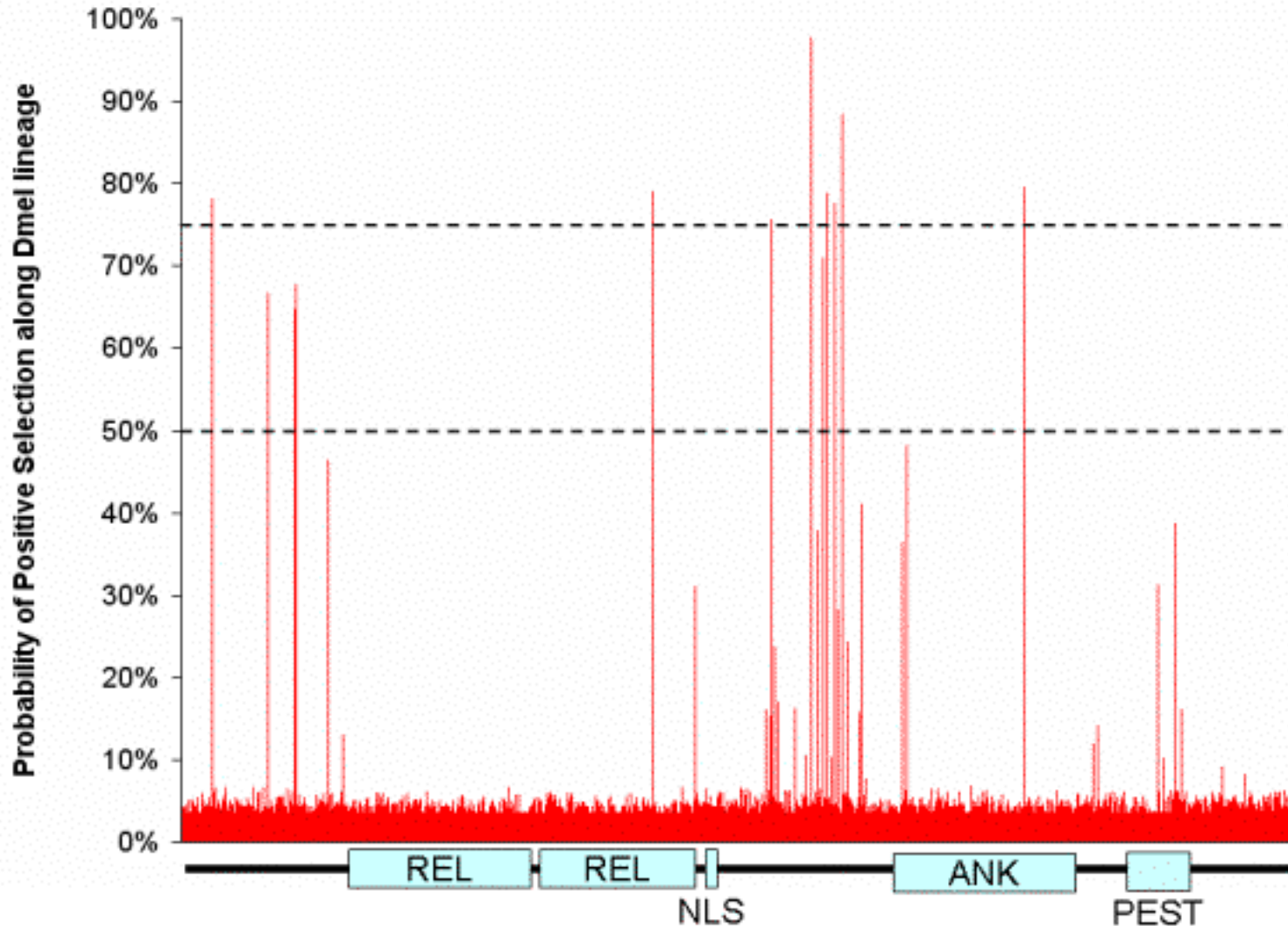
Positive selection in interactors with Relish



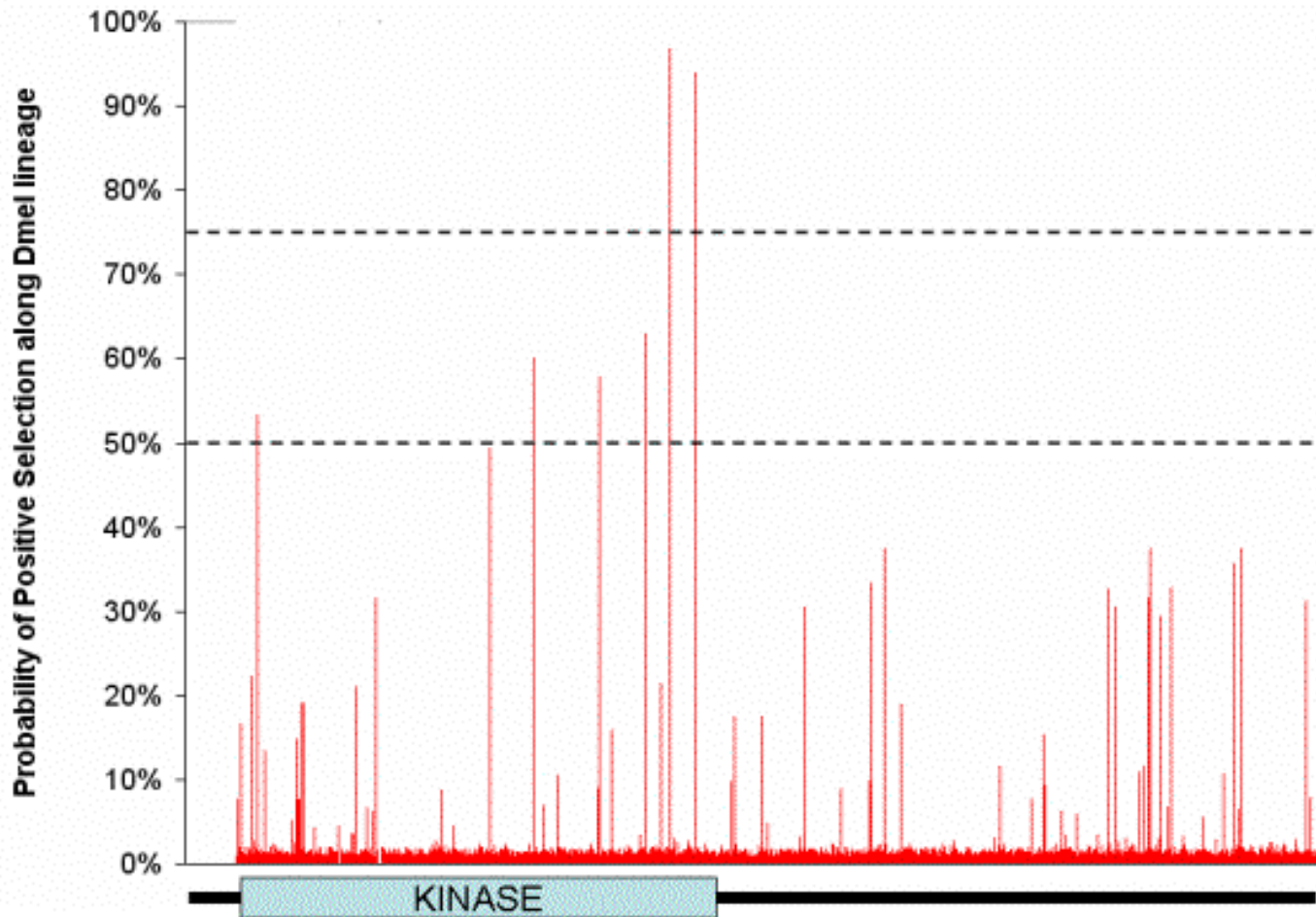
Dredd and Dnr1 appear to be positively selected across the entire phylogeny

Relish and ird5 only show evidence for positive selection in the *D. melanogaster* lineage.

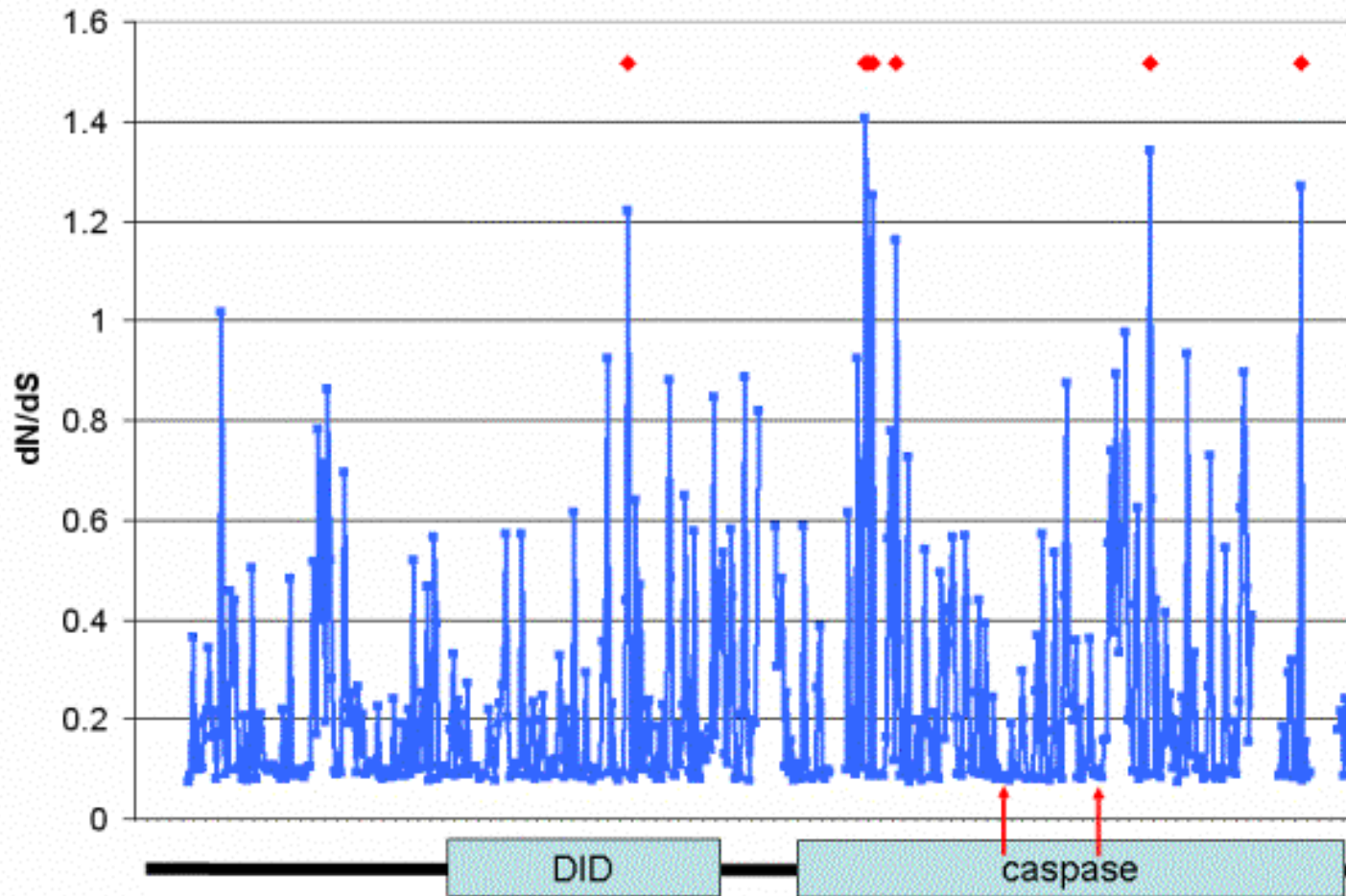
Positively selected sites in Relish are clustered in the spacer



Positively selected sites in *ird5* are clustered in the kinase domain



Clustering of positively selected sites in the caspase domain of Dredd



The clustering of positively selected sites in the interacting domains of

Dnr1, Dredd, Relish and ird5

suggests that the entire cleavage complex is evolving by positive selection in the *melanogaster* subgroup.

(probably driven by interactions with bacterial type III secretion systems)

Conclusions

- Recognition proteins display frequent gene family expansions and positive selection.
- Signaling peptides maintain one-to-one orthology. They generally show conserved sequences, but Relish appears to be a convergent evolutionary interaction, at least in insects (positively selected in termites and *D. melanogaster*).
- Rapid duplication but no positive selection in antimicrobial peptides.

Caveats

- Focus was on protein-coding genes
- Virtually ignored:
 - Regulatory evolution
 - MicroRNAs
 - Alternative splicing
 - Behavioral/ecological response to infection



Cornell

Tim Sackton
Brian Lazzaro
Todd Schlenke
Kurt McKean
Erin Hill
Sarah Stockwell
Punita Juneja
Tracy Mak
Nandita Garud

USDA Beltsville

Jay Evans

Umeå

Dan Hultmark

Berkeley

Venky Iyer
Dan Pollard

