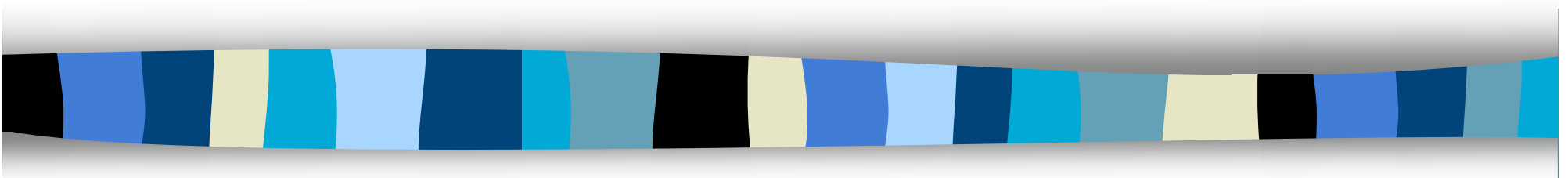


Genewise and sitewise views of adaptation: spatial variation in selection pressure in the genome

SMBE, Lyon, 7th July 2010



Daniel Wilson • University of Oxford
Molly Przeworski • University of Chicago
Peter Andolfatto • Princeton

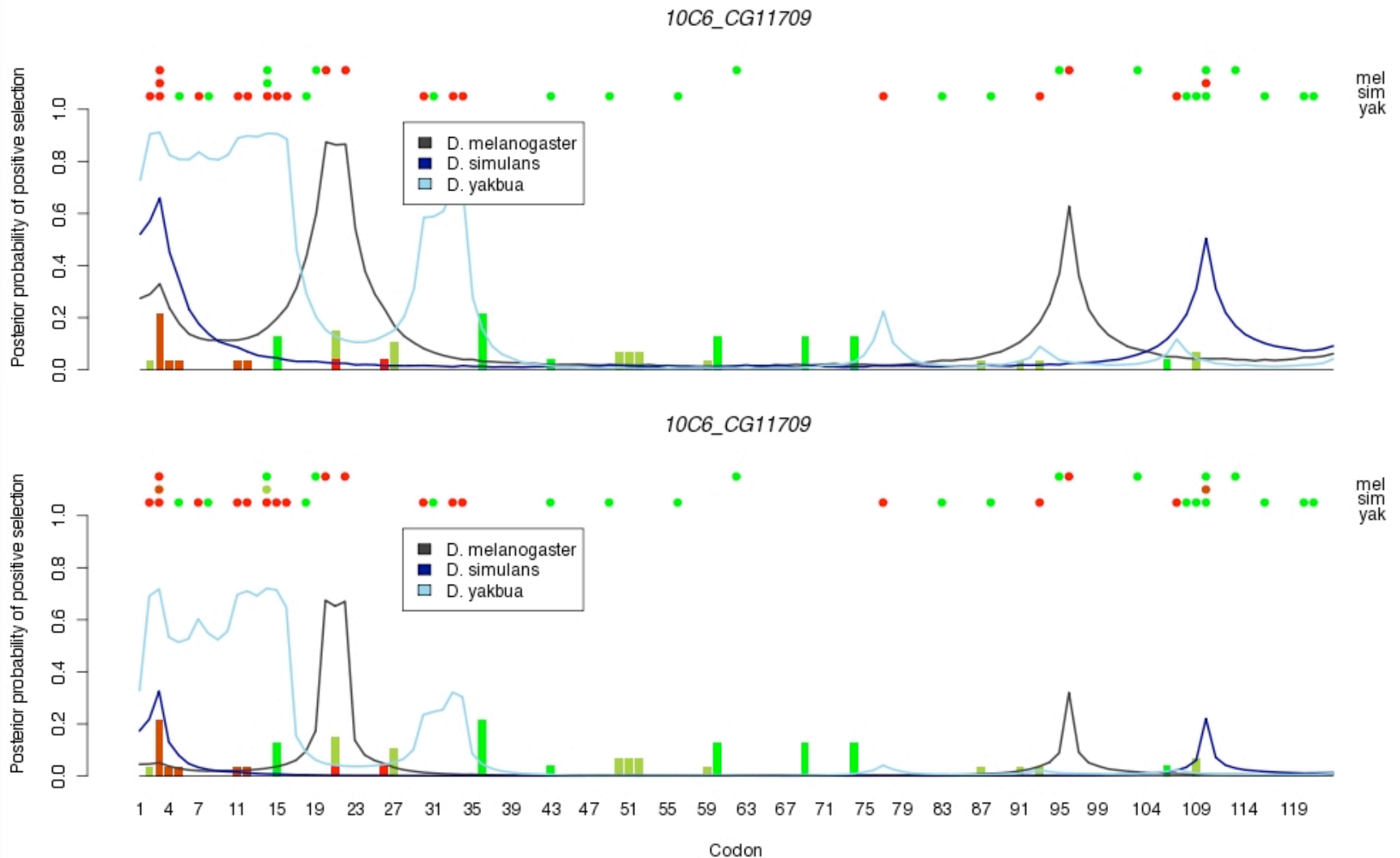




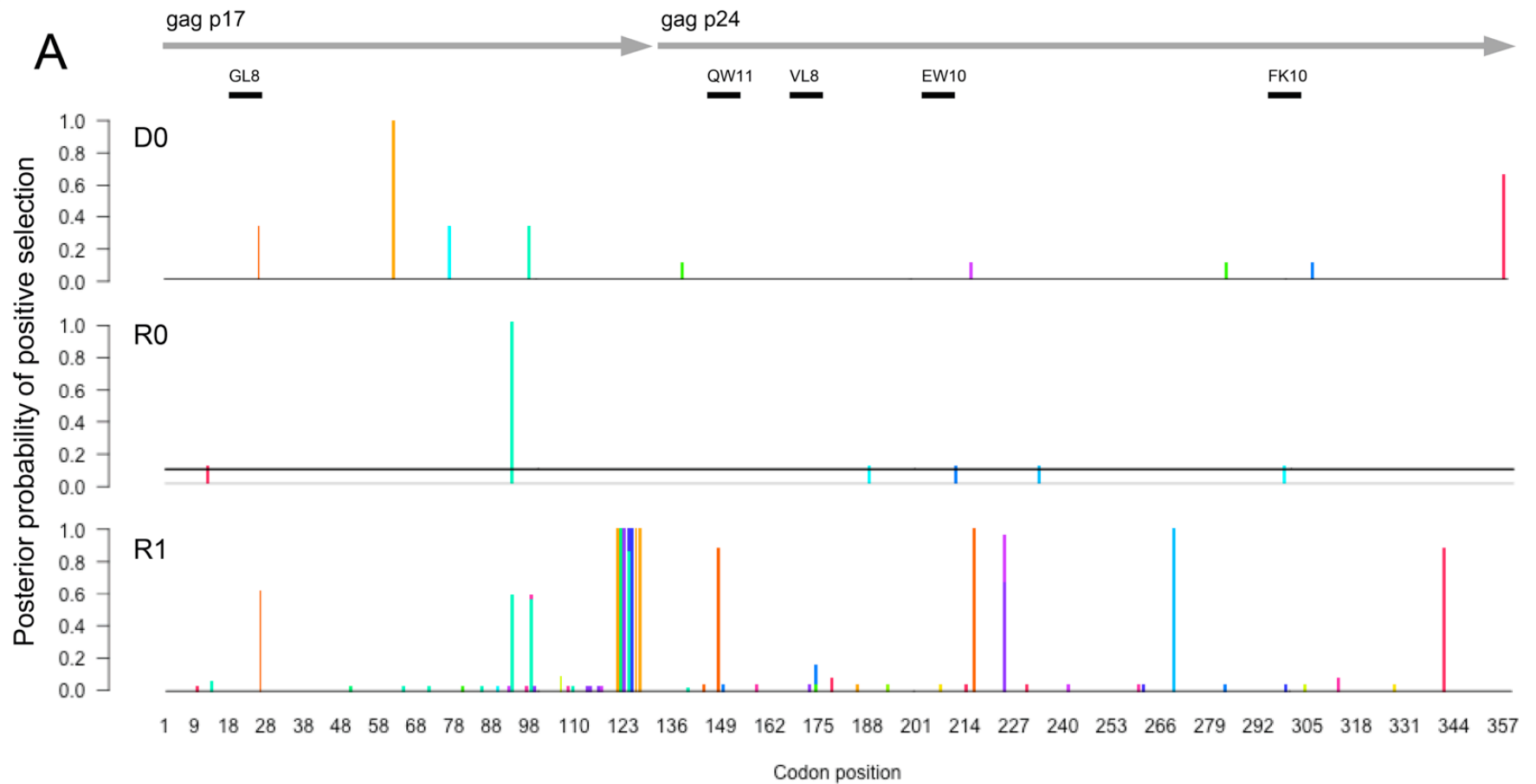
Genewise and sitewise views of adaptation: spatial variation in selection pressure in the genome

- Polymorphism within and divergence between species is informative about
 - The distribution of selective effects in the genome
 - Changes in selection intensity over time
 - The location of sites in the genome subject to adaptive substitution
- Divergence tells us about the fixed differences that distinguish species from one another.
- Polymorphism provides a snapshot of evolution at a point in time. It tells us about the mutations that did not contribute to divergence, which may be in the majority.

- The distribution of selection coefficients affects how we interpret divergence/polymorphism data at specific loci.
 - Yet we need those data before we can say anything about the distribution of selection coefficients.
- Jointly infer the nature of selection at the whole genome and sub-genomic levels.



Assumptions regarding spatial variation in selection have broad consequences.



Increased power to detecting selection through the McDonald-Kreitman test

NATURE · VOL 351 · 20 JUNE 1991

Adaptive protein evolution at the *Adh* locus in *Drosophila*

John H. McDonald & Martin Kreitman

TABLE 1 Variable nucleotides from the coding region of the *Adh* locus in *D. melanogaster*, *D. simulans* and *D. yakuba*

	Con.	<i>D. melanogaster</i>										<i>D. simulans</i>						<i>D. yakuba</i>																		
		a	b	c	d	e	f	g	h	i	j	k	l	a	b	c	d	e	f	a	b	c	d	e	f	g	h	i			j	k	l			
781	G	T	T	T	T	T	T	T	T	T	T	T	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Repl.	Fixed			
789	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Syn.	Fixed		
808	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Repl.	Fixed		
816	G	T	T	T	T	-	-	-	-	-	-	-	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	Syn.	Poly.		
834	T	-	-	-	-	-	-	-	-	-	-	-	-	C	C	-	-	-	-	C	-	-	-	-	-	-	-	-	-	-	-	-	Syn.	Poly.		
859	C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Repl.	Fixed		
867	C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Syn.	2 Poly.		
870	C	T	T	T	T	T	T	T	T	T	T	T	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Syn.	Fixed		
950	G	-	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Syn.	Poly.		
974	G	-	-	-	-	-	-	-	-	-	-	-	-	T	-	T	T	T	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Syn.	Poly.	
983	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Syn.	Fixed		
1019	C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Syn.	Poly.		
1031	C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Syn.	Poly.		
1034	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Syn.	Poly.		
1043	C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Syn.	Poly.	
1068	C	T	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Syn.	Poly.	
1089	C	-	-	-	-	-	-	-	-	-	-	-	-	A	A	A	A	A	A	A	-	-	-	-	-	-	-	-	-	-	-	-	-	Repl.	Fixed	
1101	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Repl.	Fixed	
1127	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Syn.	Fixed	
1131	C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Syn.	Poly.	
1160	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Syn.	Fixed	
1175	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Syn.	Fixed	
1178	C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Syn.	Poly.	
1184	C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Syn.	Fixed

Increased power to detecting selection through the McDonald-Kreitman test

TABLE 2 Number of replacement and synonymous substitutions for fixed differences between species and polymorphisms within species

	Fixed	Polymorphic
Replacement	7	2
Synonymous	17	42

A G -test of independence (with the Williams correction for continuity)¹ was used to test the null hypothesis, that the proportion of replacement substitutions is independent of whether the substitutions are fixed or polymorphic. $G=7.43$, $P=0.006$.

$$D_N/D_S = 1.24$$

$$P_N/P_S = 0.14$$

$$\frac{D_N / D_S}{P_N / P_S} = 8.65$$

Sawyer & Hartl's Poisson random field Parameterisation of the MK table

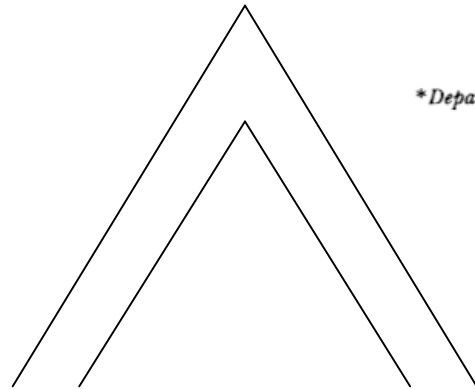
Population Genetics of Polymorphism and Divergence

Stanley A. Sawyer*[†] and Daniel L. Hartl[†]

**Department of Mathematics, Washington University, St. Louis, Missouri 63130 and [†]Department of Genetics, Washington University School of Medicine, St. Louis, Missouri 63110*

Manuscript received March 7, 1992
Accepted for publication August 12, 1992

Genetics 132: 1161–1176 (December, 1992)



melanogaster

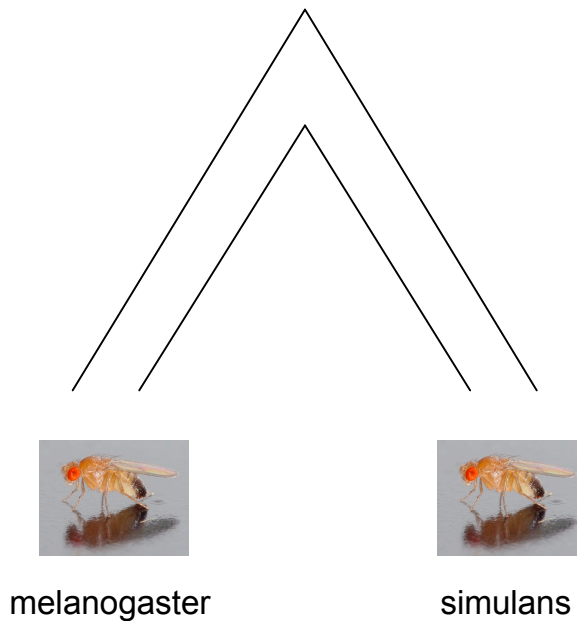


simulans

- Constant population size
- Constant mutation rate
- Constant selection intensity
- Synonymous changes are neutral
- Non-synonymous changes that are viable all have the same selection coefficient

Sawyer & Hartl's Poisson random field

Parameterisation of the MK table



- θ Population-scaled mutation rate ($4N_e\mu$)
- τ Divergence time
- γ Population-scaled selection coefficient ($4N_e s$)
- f Proportion nonsynonymous changes that are inviable

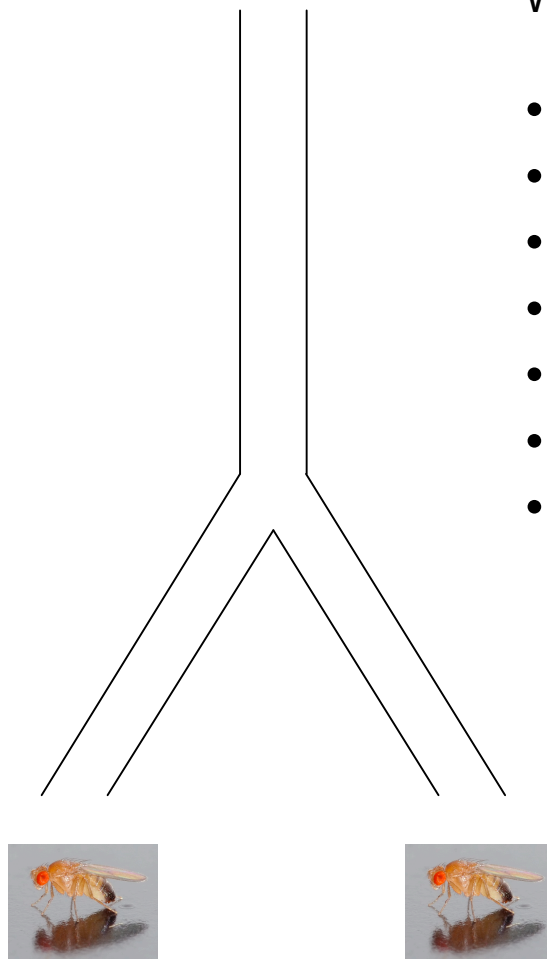
	Fixed	Polymorphic
Replacement	7	2
Synonymous	17	42

$$\theta = 32 \quad \tau = 1.9 \quad \gamma = 8.4 \quad f = 0.99$$

Multispecies codon-based model



yakuba



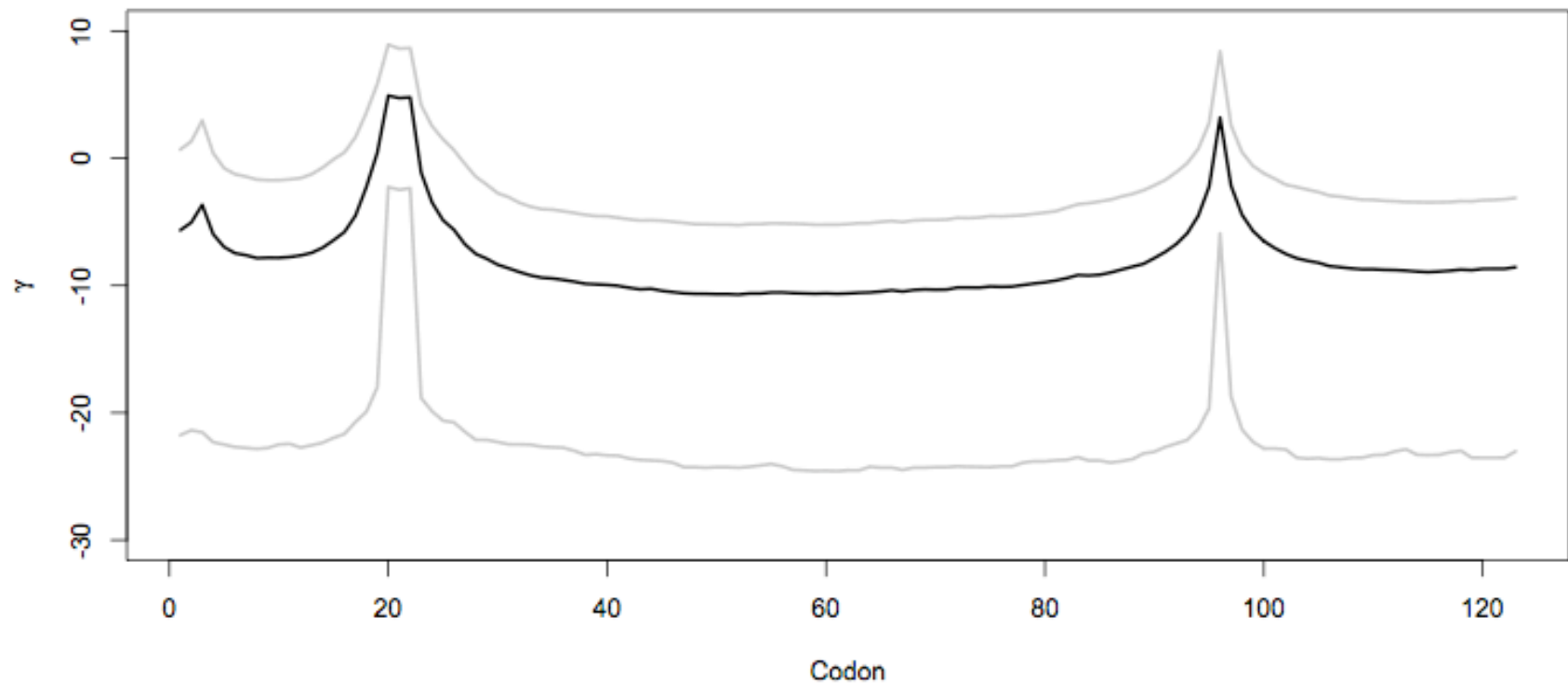
melanogaster

simulans

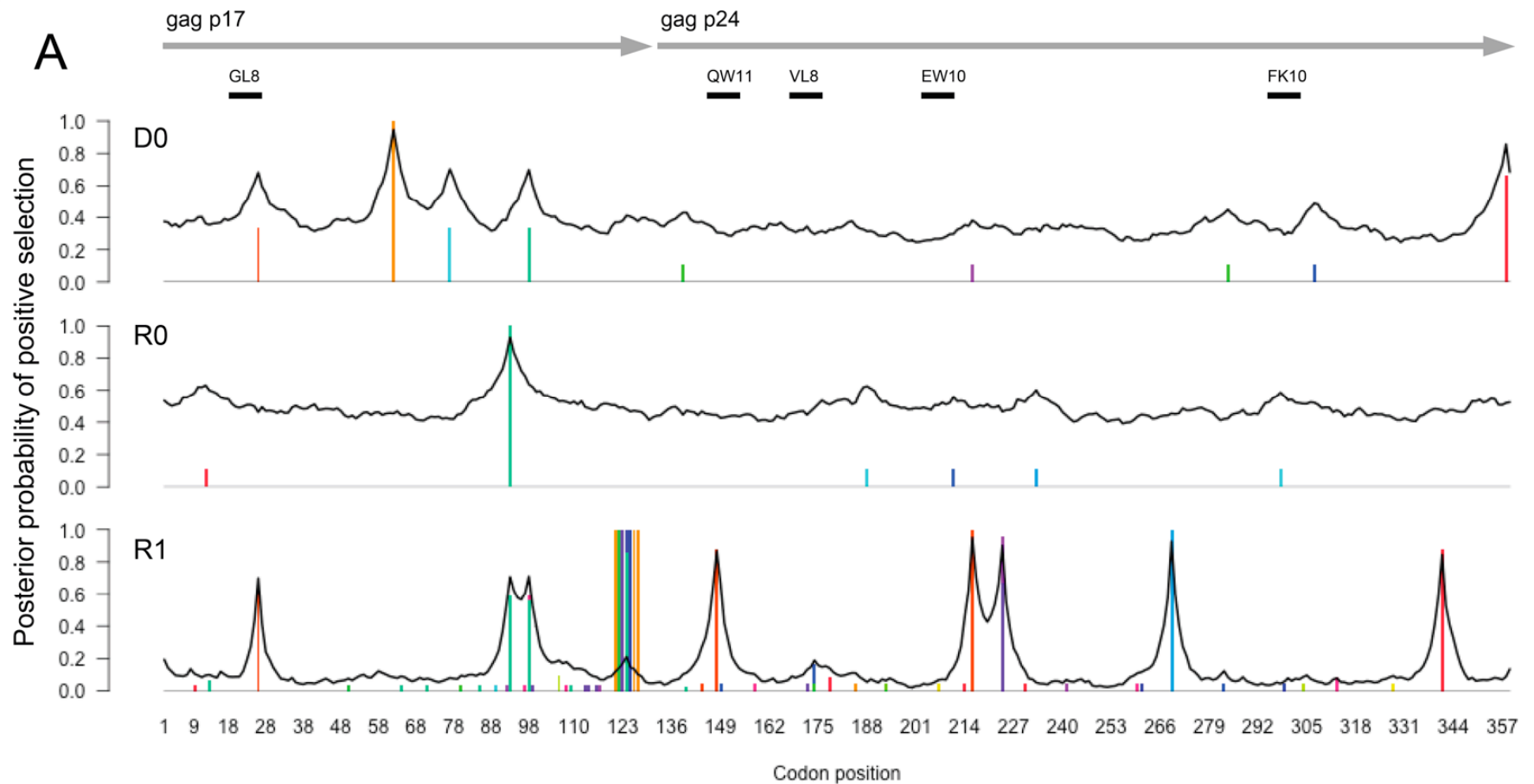
Wilson and Przeworski, in preparation

- Arbitrarily many species
- Explicit codon-based model
- Transition-transversion ratio
- Multiple alleles at the same site
- Probabilistic inference of ancestral states
- Different parameters for each lineage
- Bayesian

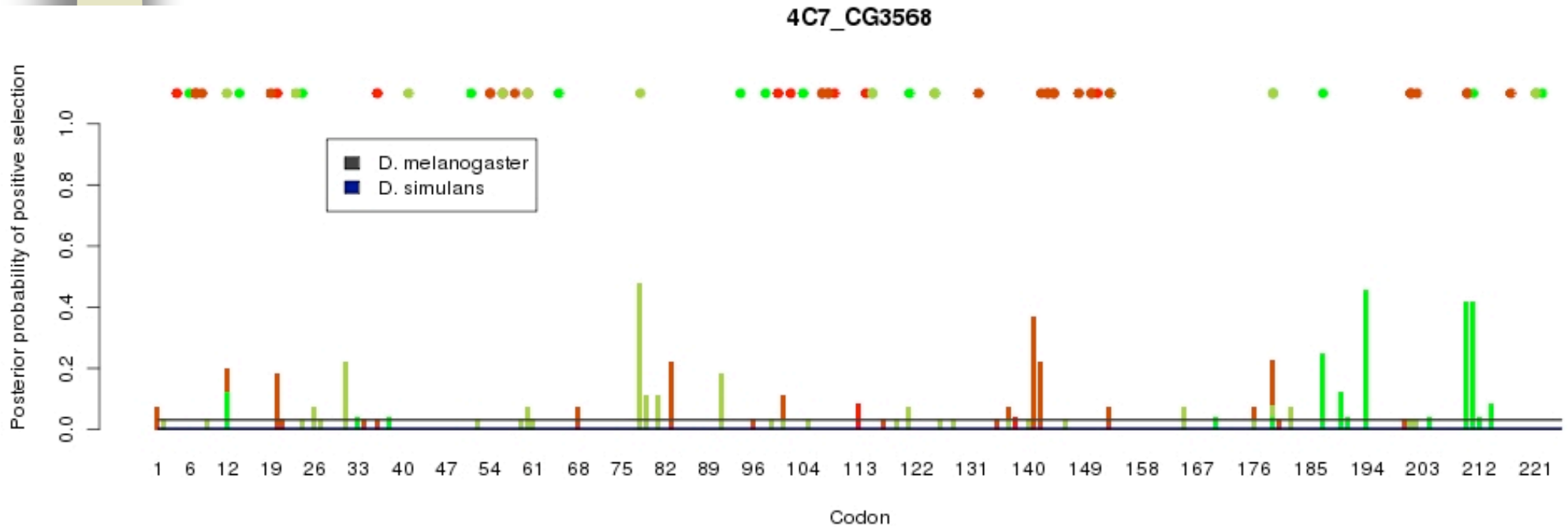
Sliding window model of variation in selection pressure



Assumptions regarding spatial variation in selection have broad consequences.



Analysis of the *Drosophila* X chromosome



Key

melanogaster

simulans



Fixed difference: non-synonymous



Fixed difference: synonymous



Polymorphism: non-synonymous



Polymorphism: synonymous

Analysis of the *Drosophila X* chromosome



melanogaster

●	17
●	12
■	2
■	14

simulans

●	9
●	3
■	20
■	26

Fixed difference: non-synonymous

Fixed difference: synonymous

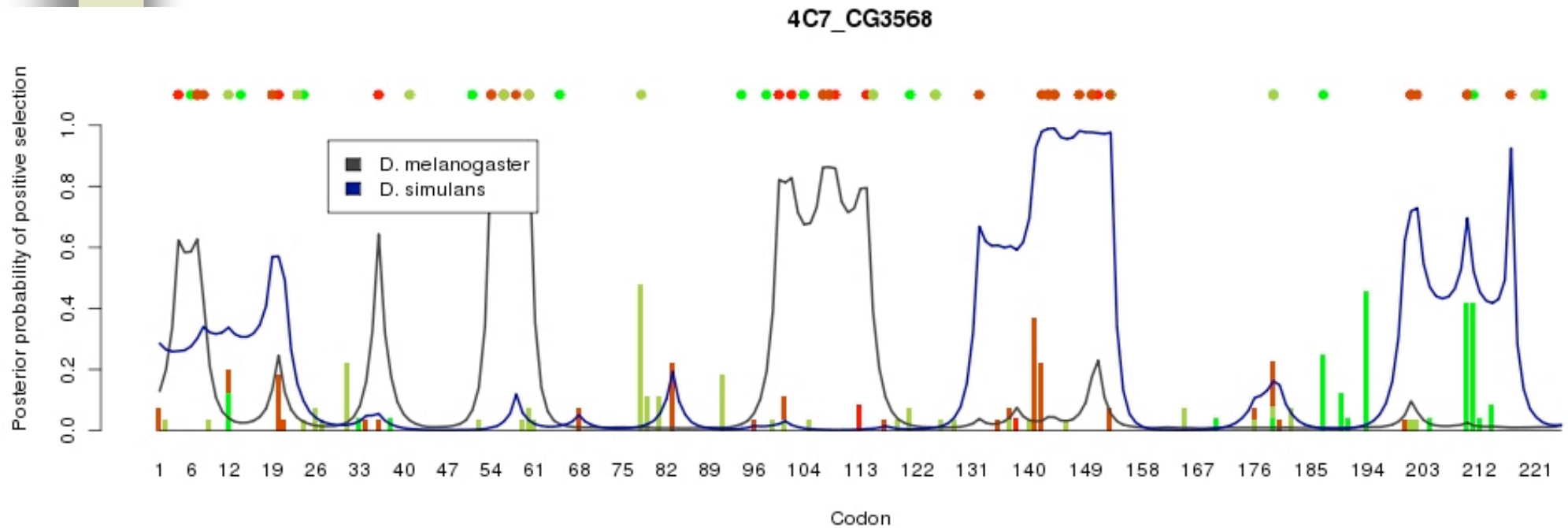
Polymorphism: non-synonymous

Polymorphism: synonymous

$$\frac{D_N / D_S}{P_N / P_S} = \frac{0.47}{0.05}$$
$$= 9.9$$
$$p = 0.01$$

$$\frac{D_N / D_S}{P_N / P_S} = \frac{1.0}{0.25}$$
$$= 3.9$$
$$p = 0.11$$

Analysis of the *Drosophila* X chromosome



Key

melanogaster

simulans



Fixed difference: non-synonymous



Fixed difference: synonymous

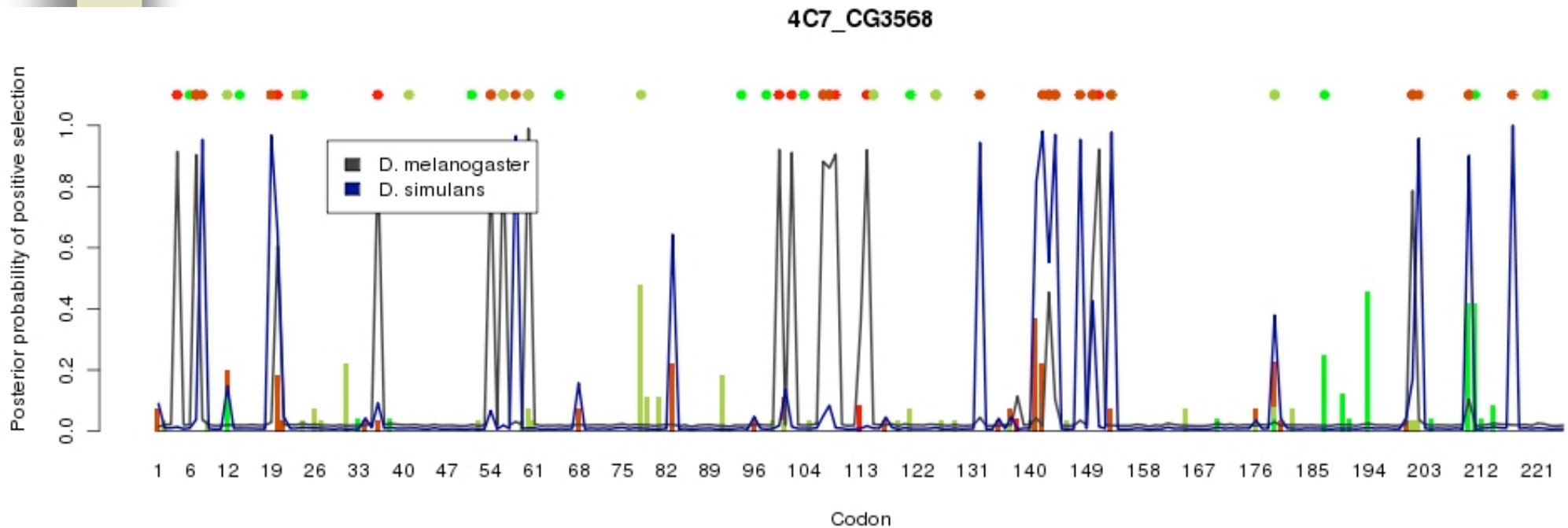


Polymorphism: non-synonymous



Polymorphism: synonymous

Analysis of the *Drosophila* X chromosome



Key

melanogaster

simulans



Fixed difference: non-synonymous



Fixed difference: synonymous



Polymorphism: non-synonymous



Polymorphism: synonymous



Mean selection coefficient

	<i>D. melanogaster</i>	<i>D. simulans</i>
Genewise	-139.4	-156.2
Sliding window	-332.0	-321.4
Sitewise	-265.0	-253.8

Percent genome adaptively evolving

	<i>D. melanogaster</i>	<i>D. simulans</i>
Genewise	7.1	10.6
Sliding window	5.0	6.5
Sitewise	4.0	3.1



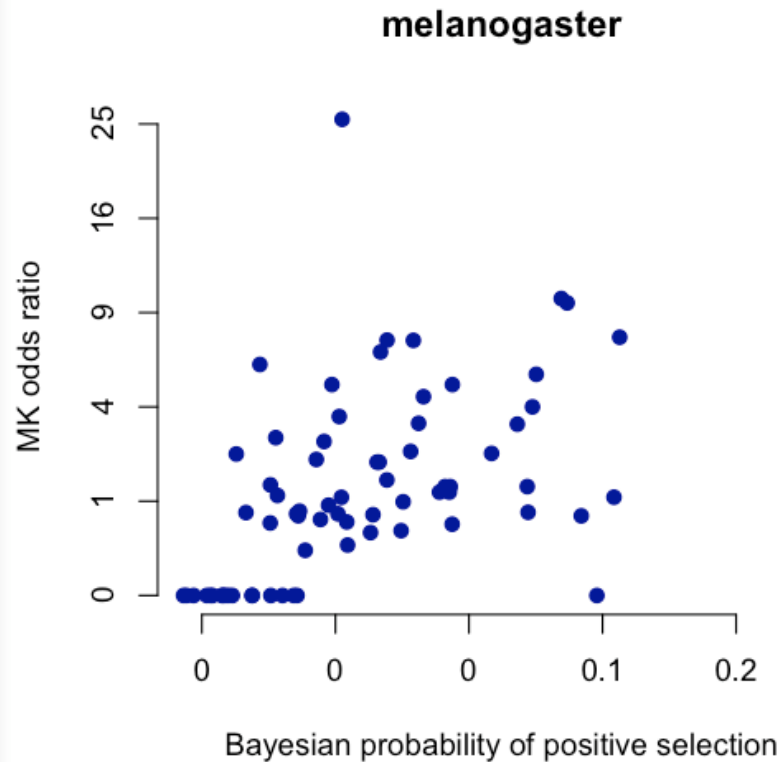
Mean window length (codons)

	<i>D. melanogaster</i>	<i>D. simulans</i>
Genewise	Whole gene	Whole gene
Sliding window	3.83	9.33
Sitewise	1	1

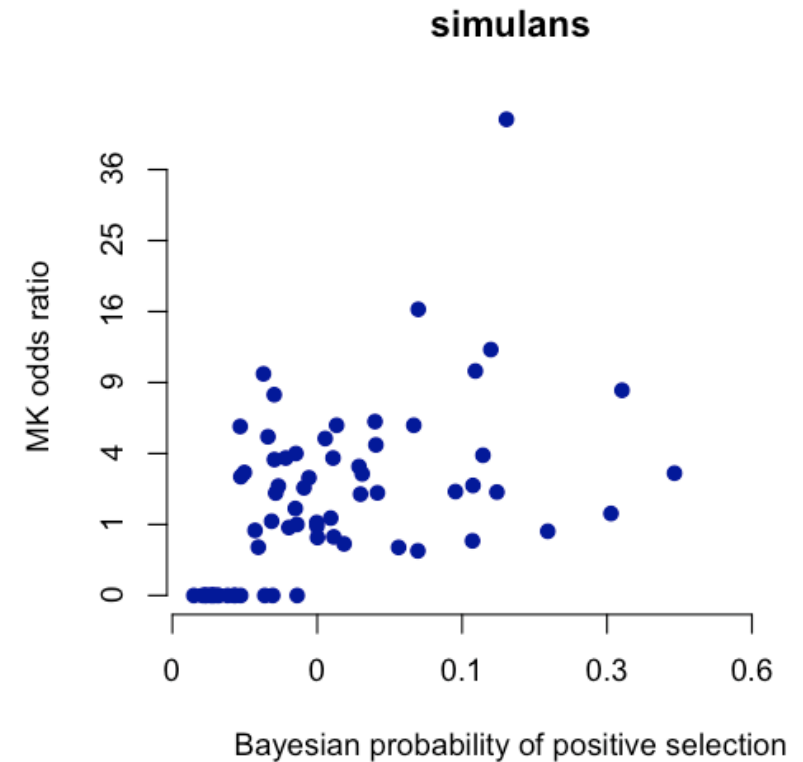
Branch length to common ancestor

	<i>D. melanogaster</i>	<i>D. simulans</i>
Genewise	0.196	0.072
Sliding window	0.162	0.070
Sitewise	0.158	0.064

Weak congruence between sitewise and genewise detection of positive selection



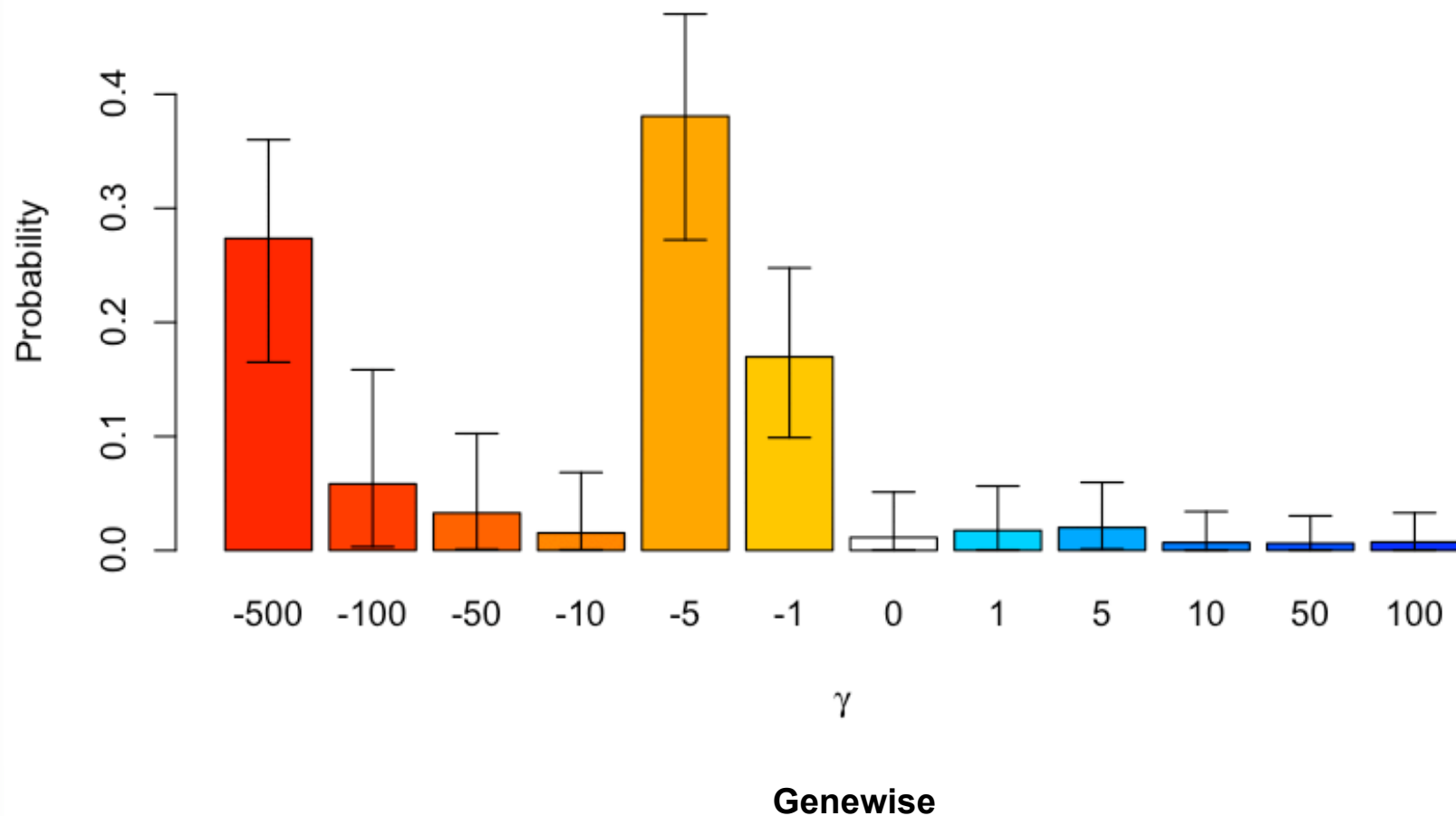
$$R^2 = 0.22$$



$$R^2 = 0.26$$

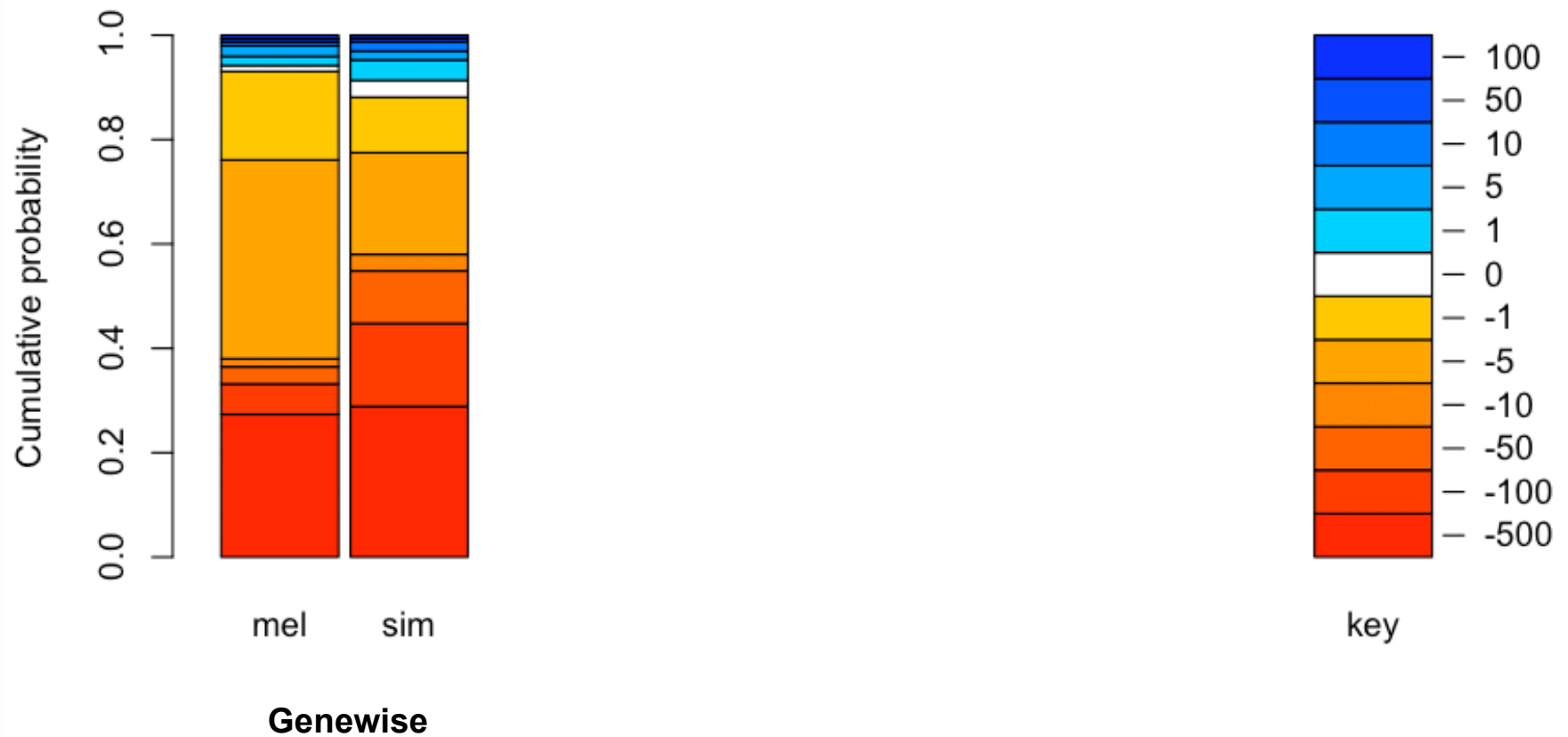
How the inferred distribution of selection coefficients is affected

Distribution of selection coefficients: mutations

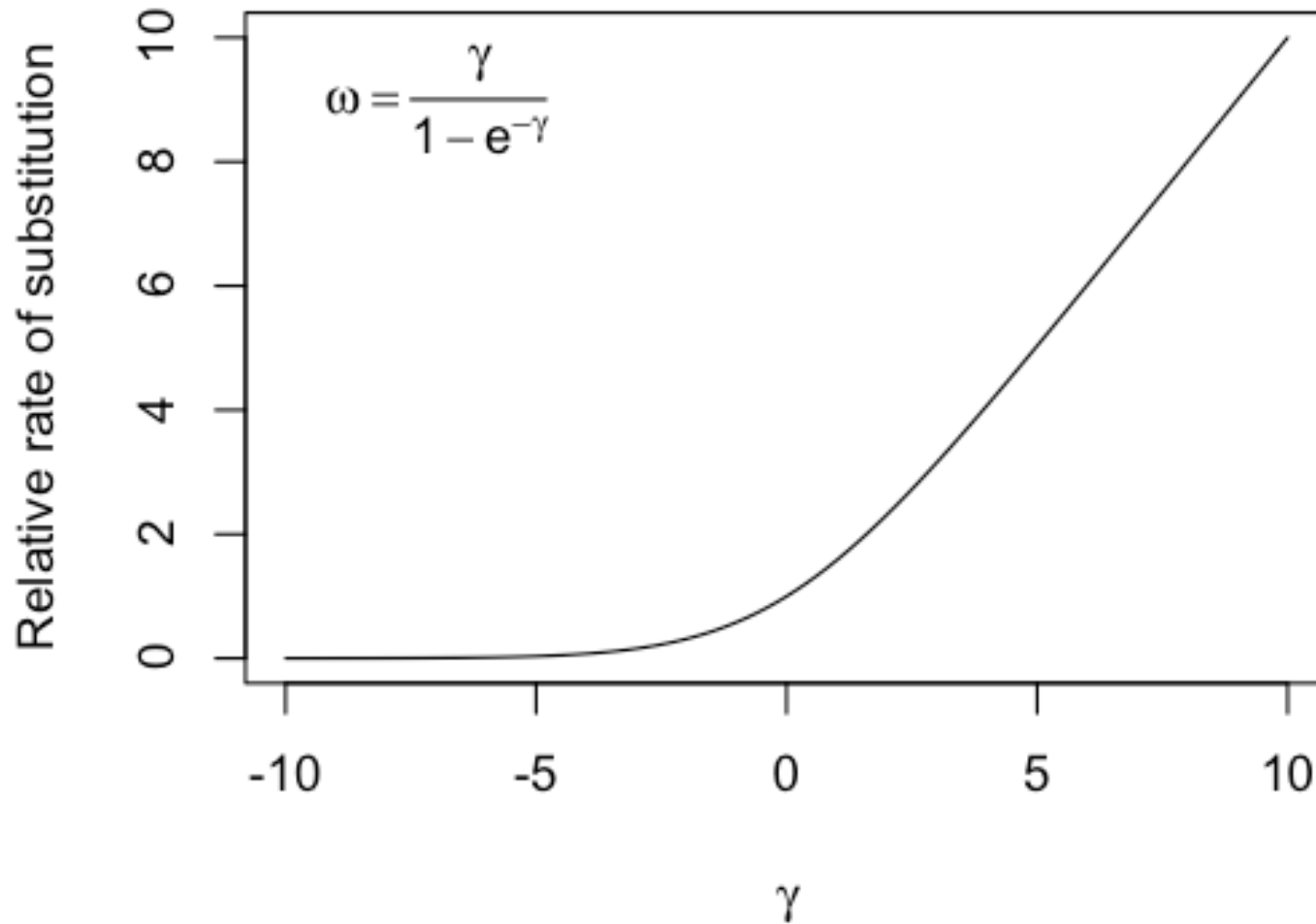


How the inferred distribution of selection coefficients is affected

Distribution of selection coefficients: mutations

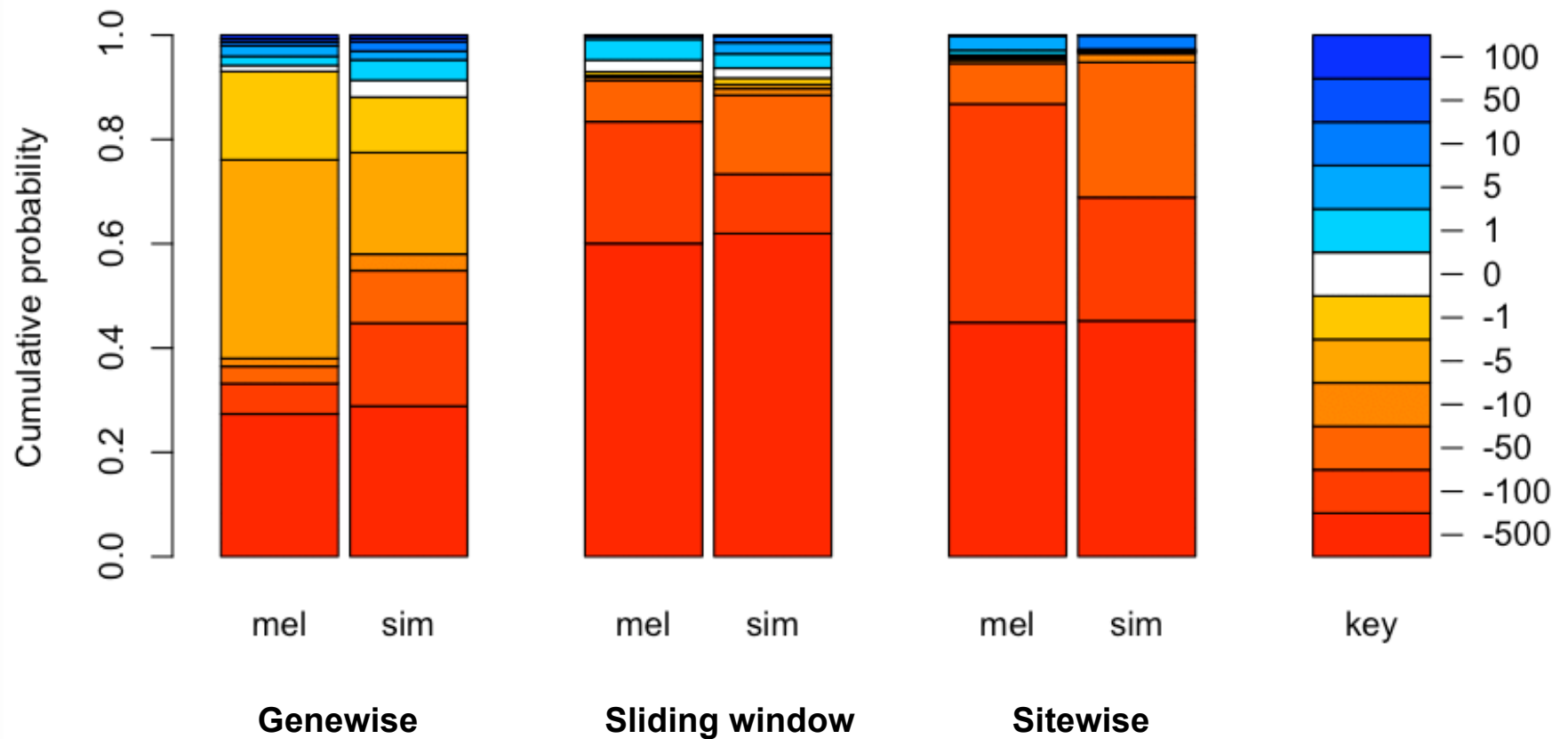


Beneficial alleles are greatly over-represented in the mutations that fix



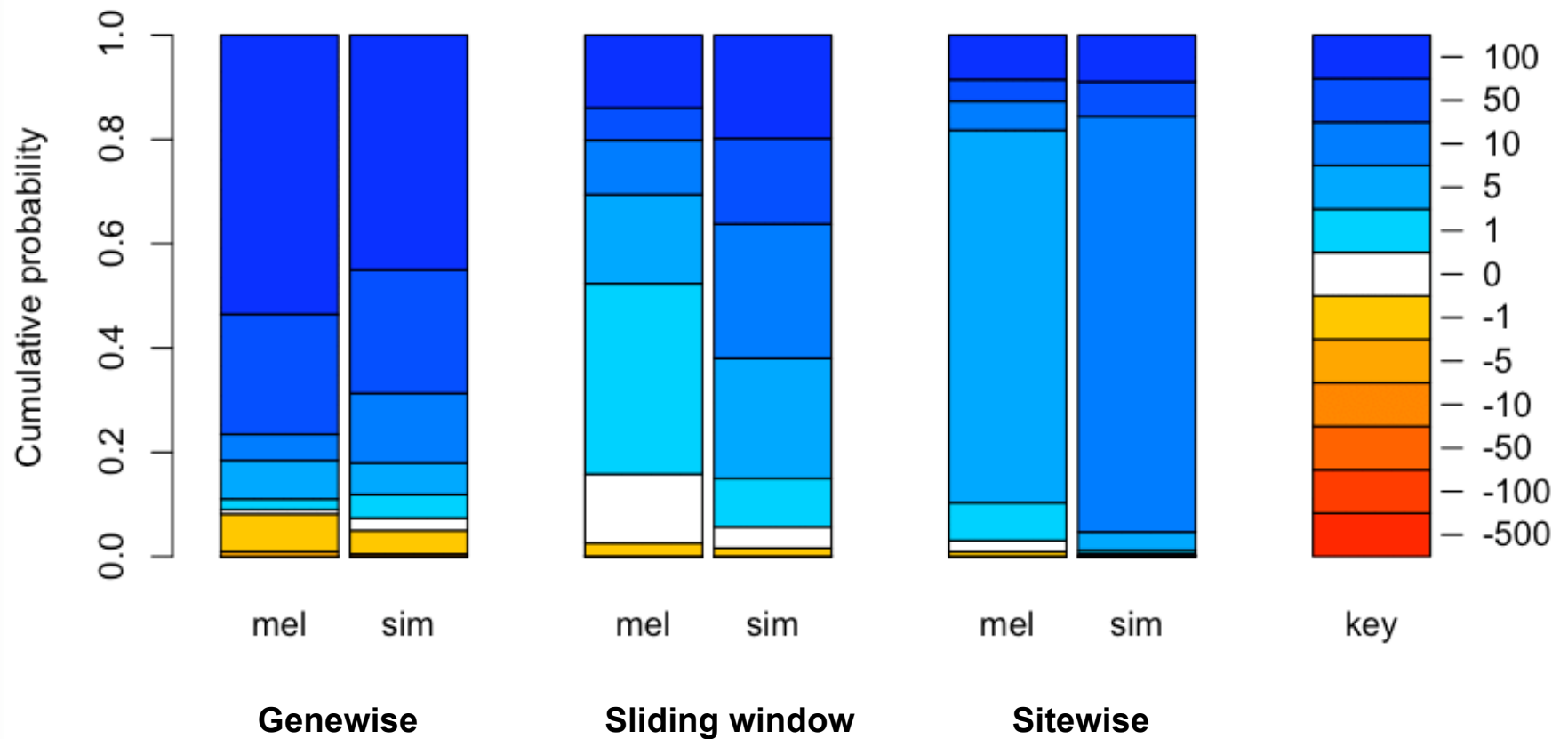
The distribution of selection coefficients for new mutants

Distribution of selection coefficients: mutations



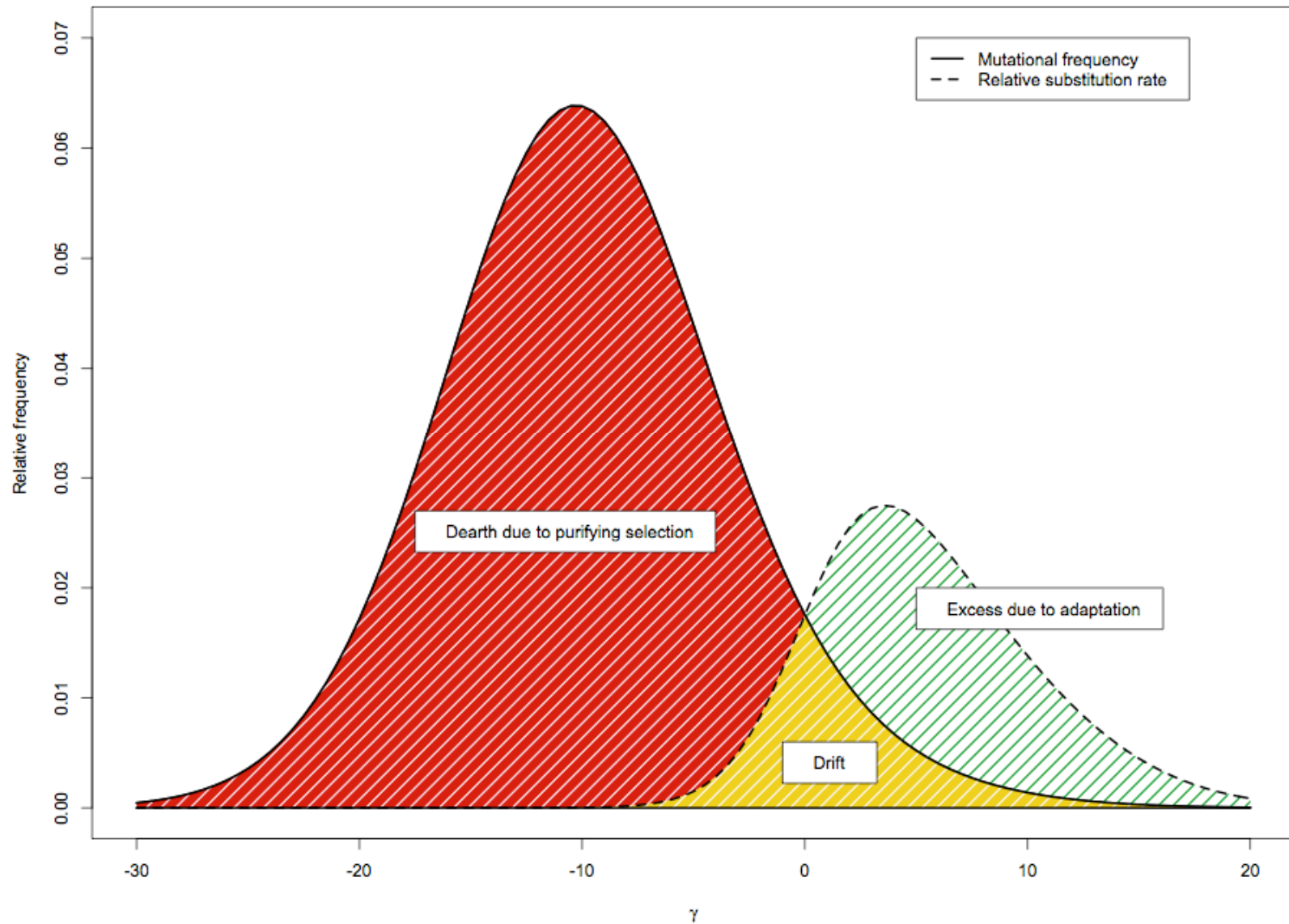
The distribution of selection coefficients for amino acid substitutions

Distribution of selection coefficients: substitutions



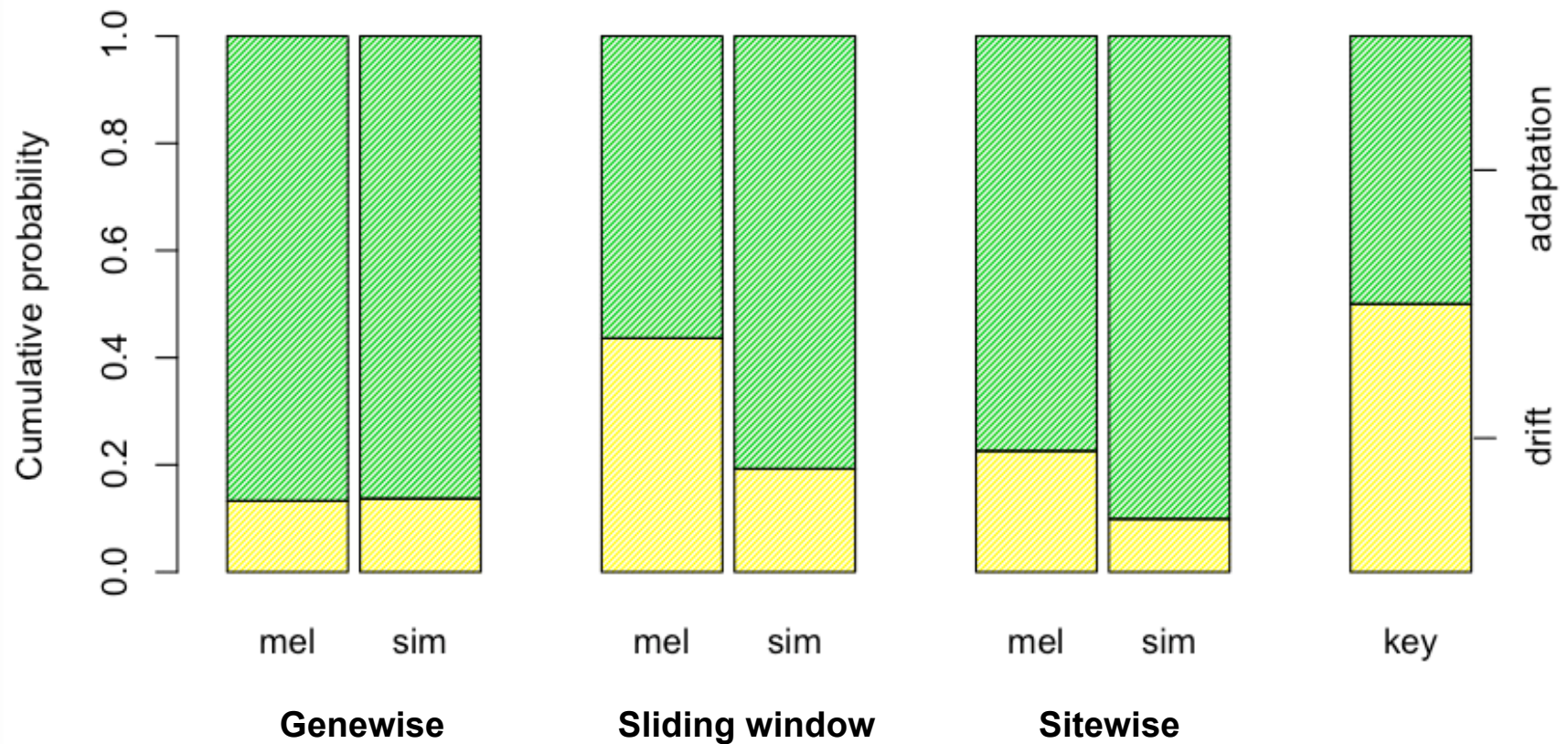
Proportion of fixations driven by adaptation

Rate of substitution relative to mutational frequency



Proportion of amino acid substitutions attributable to adaptation

Contribution to substitution of drift versus adaptation

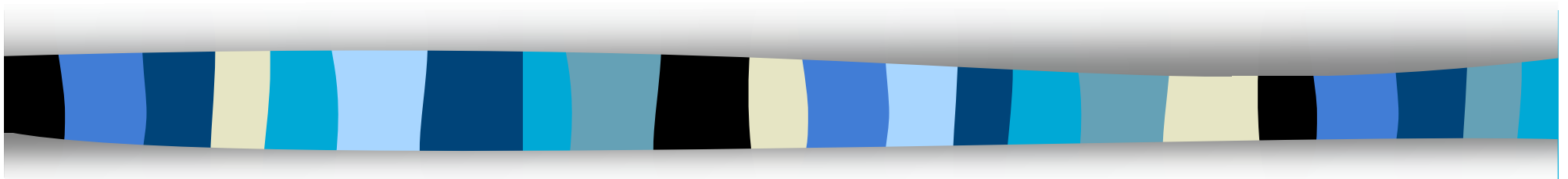


In summary



- Models of natural selection allow you to
 - Quantify the distribution of selection coefficients
 - Visualize changes in the selective regime between species
 - Localize the signal of selection
- The model of spatial variation in selection pressures within the genome affects both
 - The estimated distribution of fitness effects
 - The evidence for selection at an individual site
 - Conclusions arising
- In order to understand the effect of selection more clearly, we need to appreciate the nature of fine-scale variation within the genome.
 - Combined analyses of polymorphism and divergence allow us to do this.

www.danielwilson.me.uk



daniel.wilson@ndm.ox.ac.uk