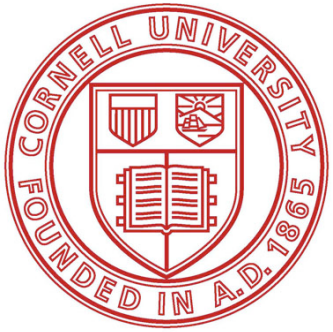
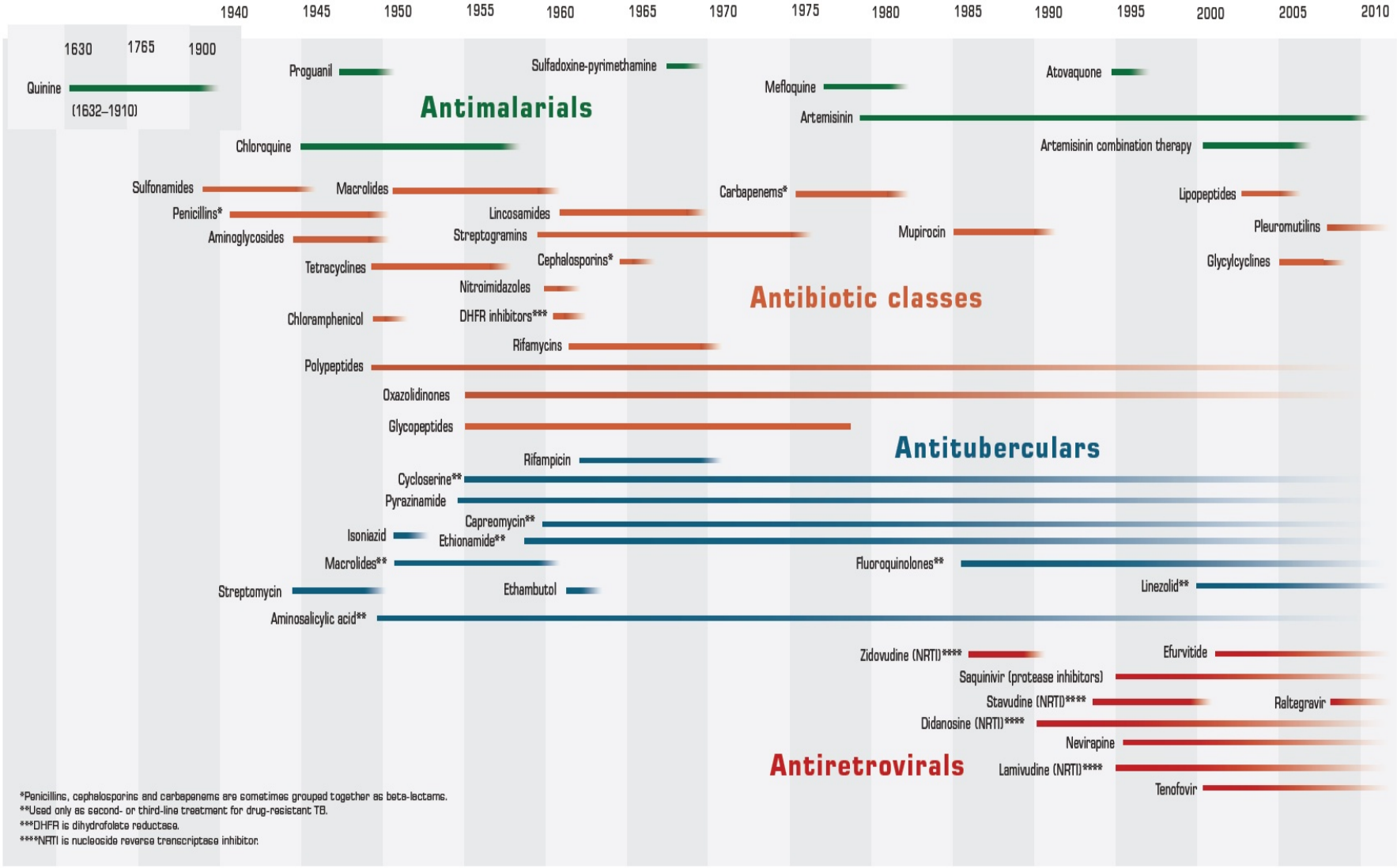
Population genetics of selective sweeps in drug resistance evolution

Philipp W. Messer

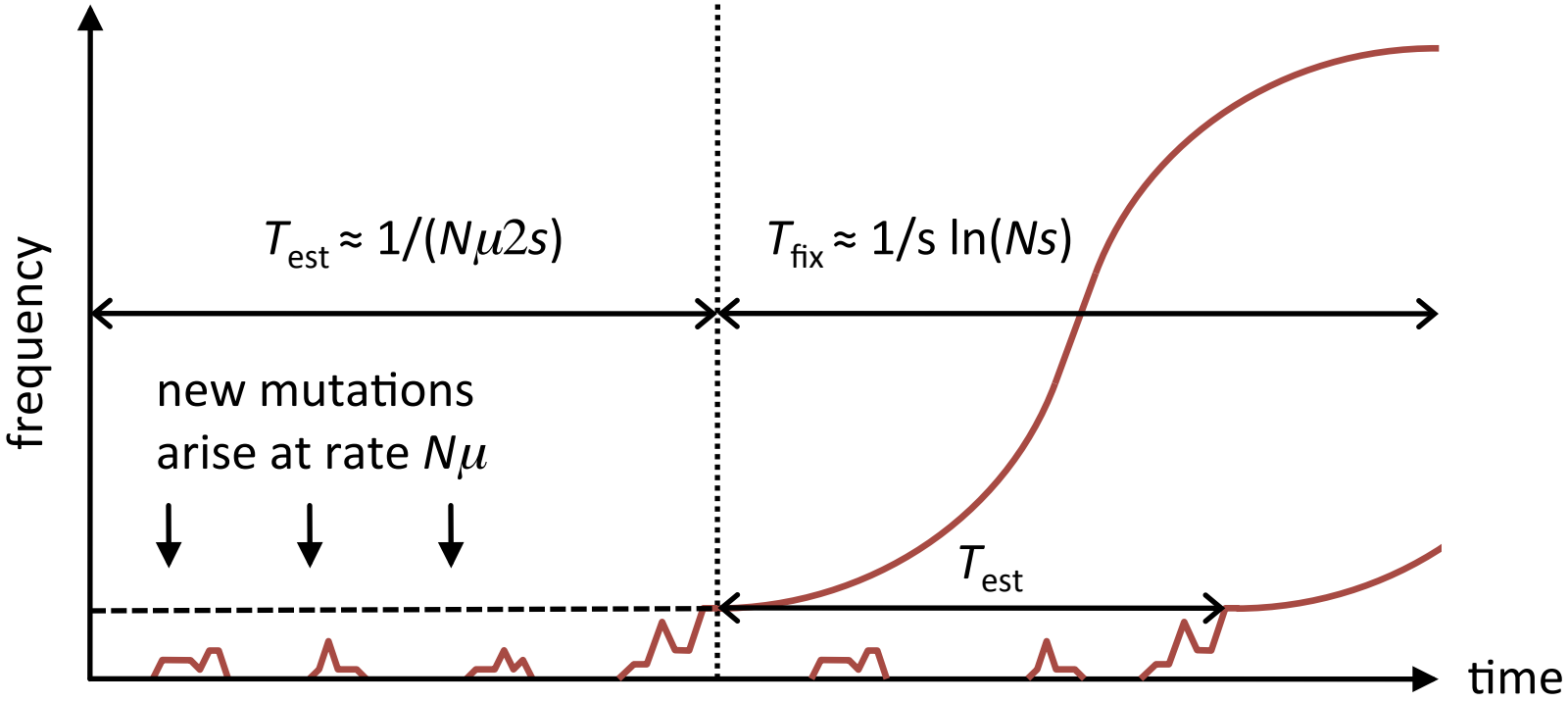


Dept. of Biological Statistics and Computational Biology, Cornell



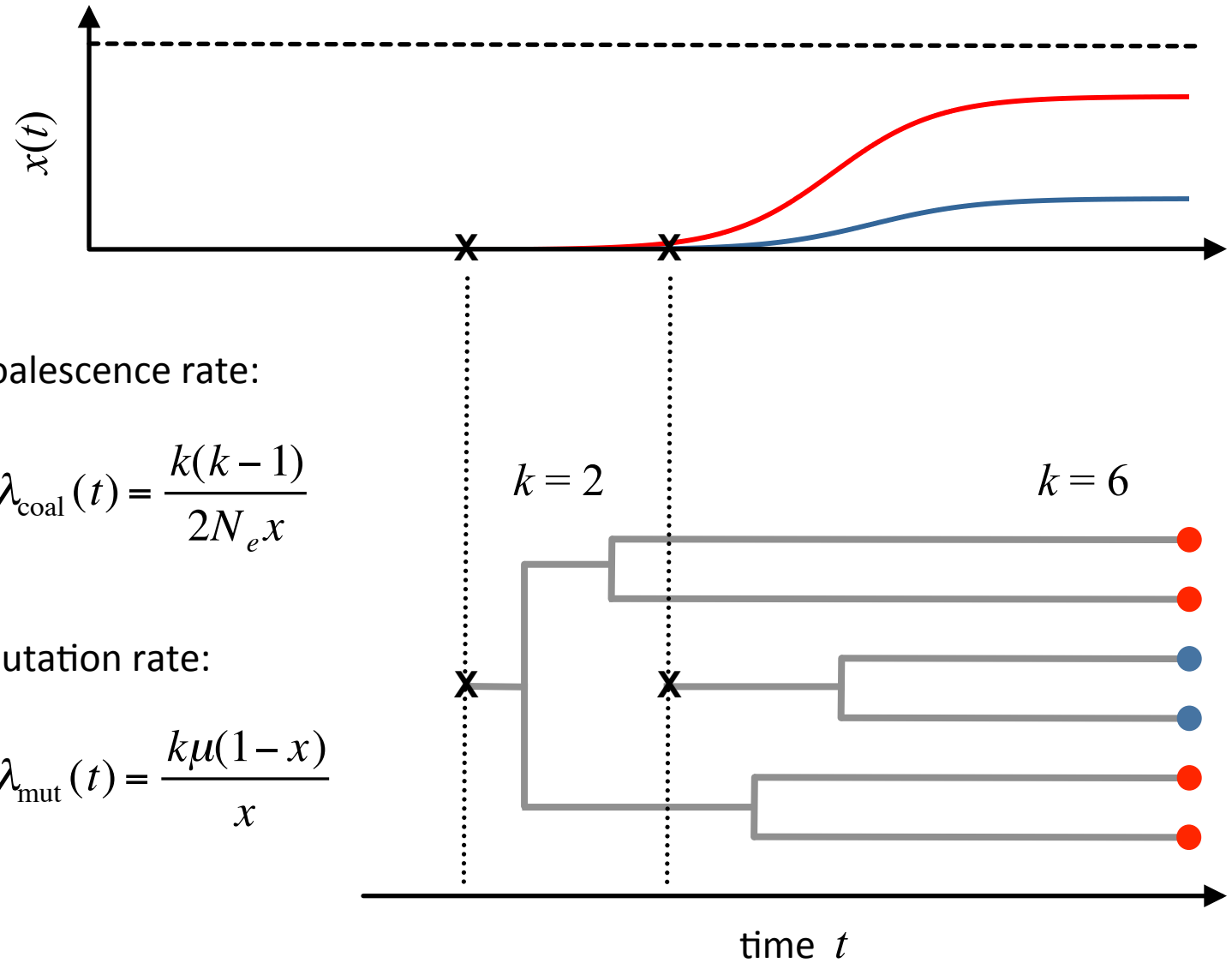
Nugent et al, *CDG Reports* (2010)

Adaptation from *de novo* mutation



recurrent *de novo* origins likely when: $T_{est} < T_{fix} \Rightarrow 2N\mu > 1/\ln(Ns)$

Coalescent with “killings” framework



coalescence rate:

$$\lambda_{\text{coal}}(t) = \frac{k(k-1)}{2N_e x}$$

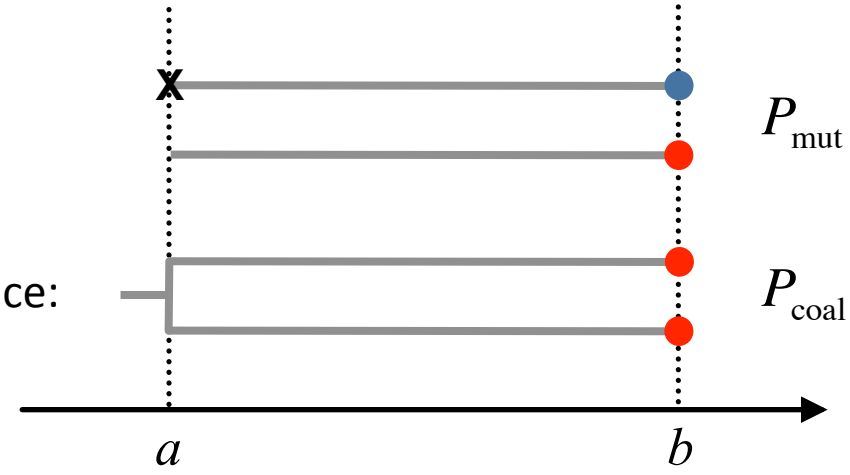
mutation rate:

$$\lambda_{\text{mut}}(t) = \frac{k\mu(1-x)}{x}$$

When do we expect soft sweeps?

1) mutation:

2) coalescence:



$$P_{\text{mut}}(a,b) = \int_a^b \lambda_{\text{mut}}(t)\, dt$$

$$P_{\text{coal}}(a,b) = \int_a^b \lambda_{\text{coal}}(t)\, dt$$

$$P_{\text{soft}} = \frac{P_{\text{mut}}(a,b)}{P_{\text{mut}}(a,b) + P_{\text{coal}}(a,b)} = \frac{\Theta}{1+\Theta}$$

idealized WF population of constant size: $\Theta = 2\mu N$

$$P_{\text{soft}} \approx \begin{cases} \Theta & \text{if} \quad \Theta << 1 \\ 1/2 & \text{if} \quad \Theta = 1 \\ 1 & \text{if} \quad \Theta >> 1 \end{cases}$$

1. independent of strength of selection

2. weak dependence on sample size

Soft sweeps and demography

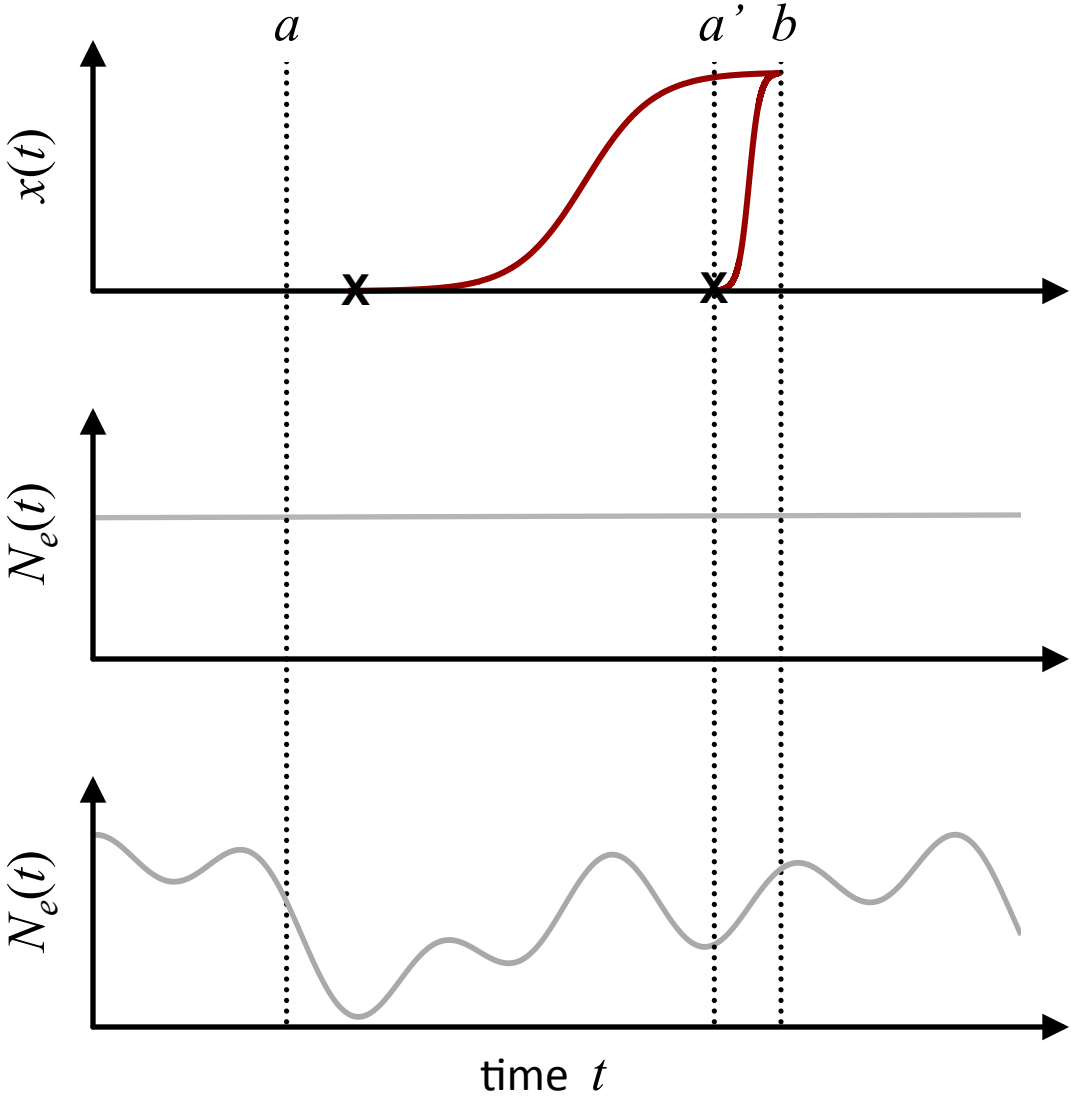
$$P_{soft} = \frac{\Theta}{1+\Theta}$$

constant $N(t)$:

$$\Theta = 2\mu N_e(a)$$

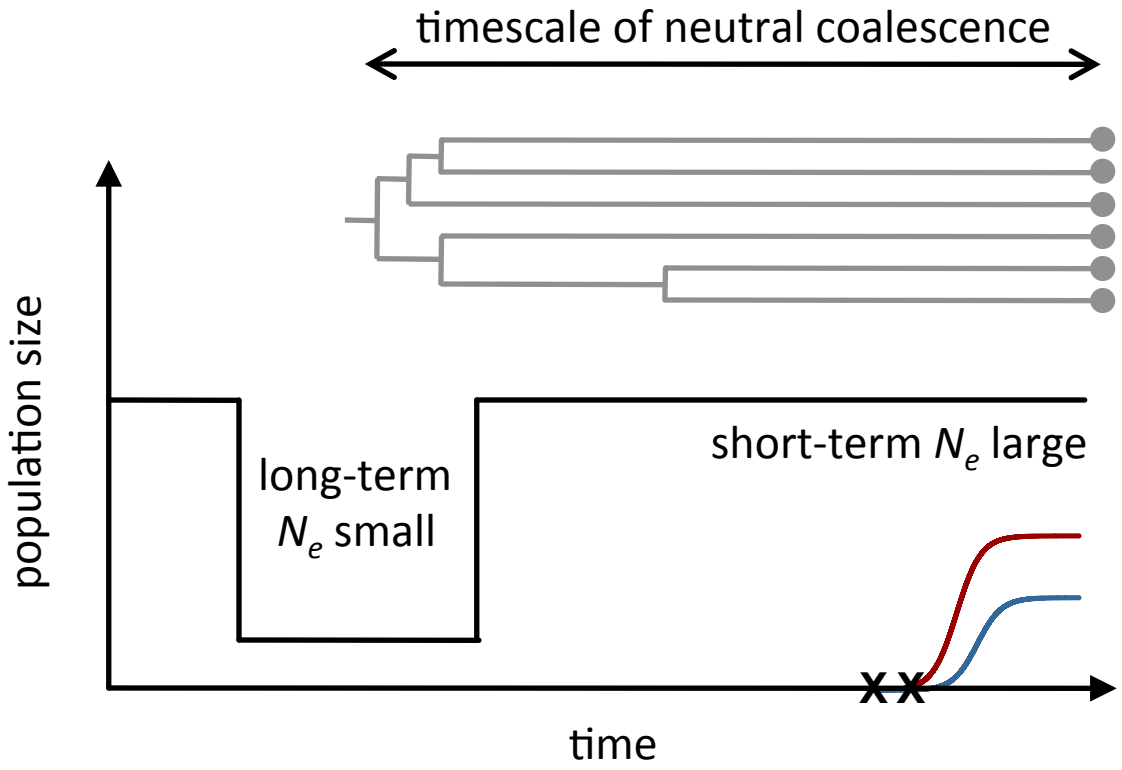
fluctuating $N(t)$:

$$\Theta = 2\mu \langle N_e \rangle_{\text{harmonic}(a,b)}$$



Wilson, Petrov and Messer, *Genetics* (2014)

Short/long-term effective population size



Recent adaptation depends primarily on short-term N_e

Messer and Petrov, *Trends Ecol Evol* (2013)

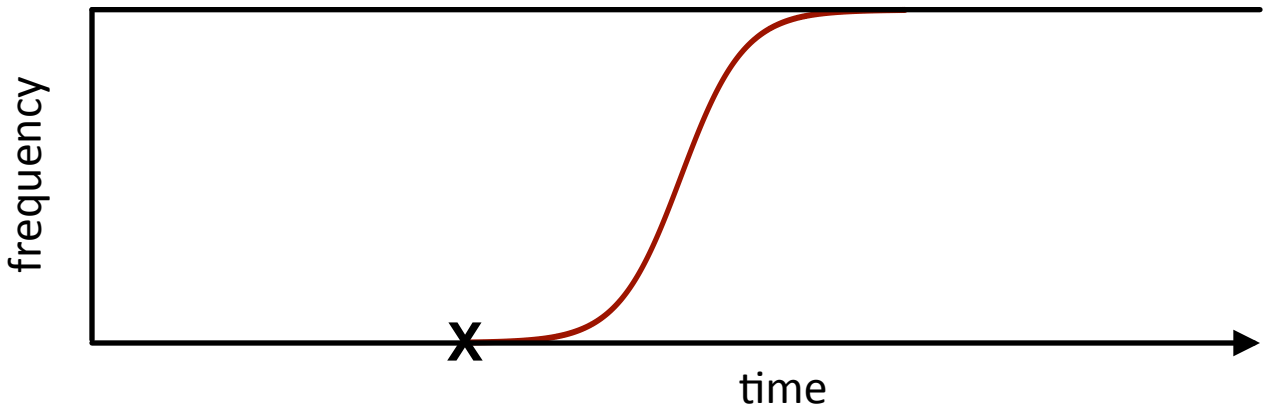
Predictions from population genetic theory

Short-term N_e relevant for rapid adaptation can be much larger than diversity-based estimates

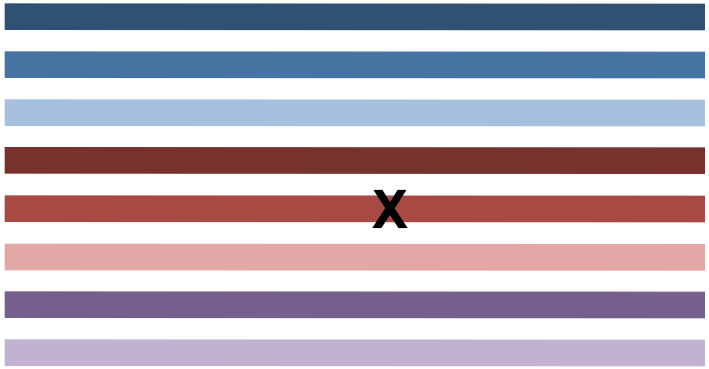
Rapid adaptation should commonly produce soft sweeps from recurrent *de novo* mutations

Can we identify the signatures of these soft sweeps in population genomic data?

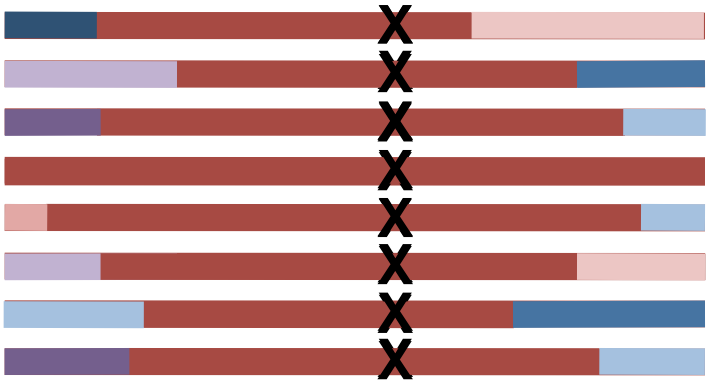
Selective sweeps



before:

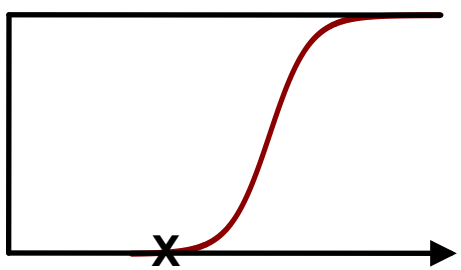


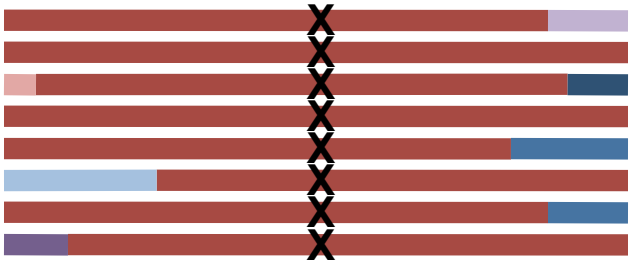
after:



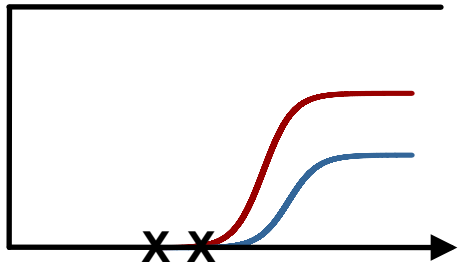
Hard and soft selective sweeps

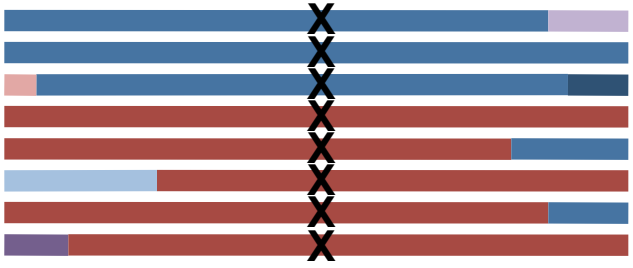
Hard sweep:





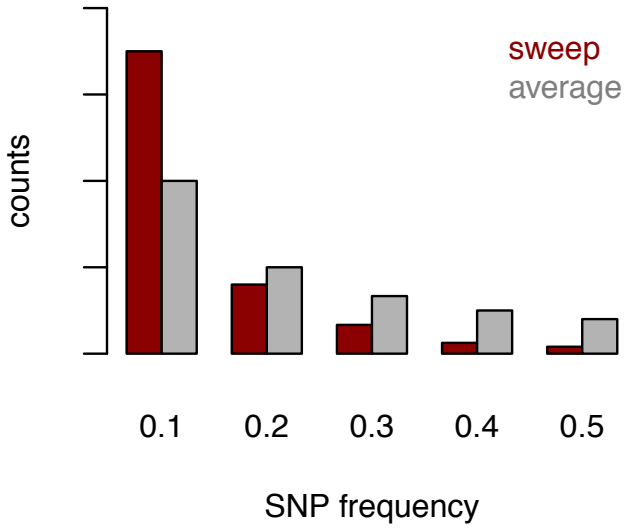
Soft sweep:





Genomic scans for selective sweeps

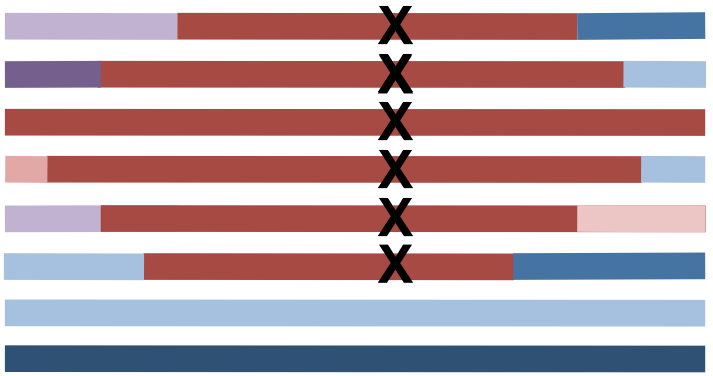
Site frequency spectrum (SFS) based methods:



TD: Tajima, *Genetics* 1989

CLR: Nielsen et al, *Genome Res* 2005

Haplotype homozygosity based methods:



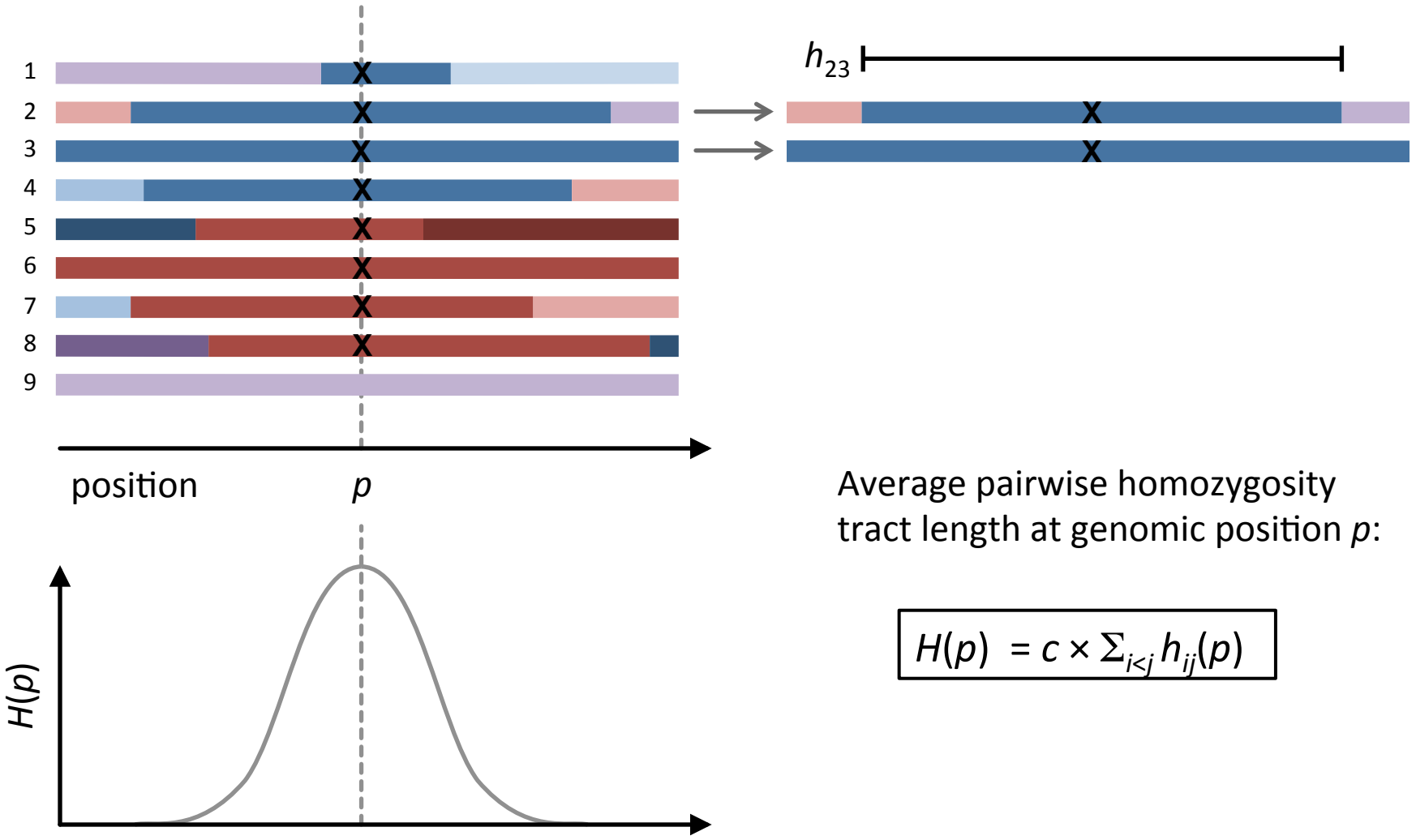
iHS : Voight et al, *PLoS Biology* 2005

nSL: Ferrer-Admetlla et al, *MBE* 2014

H12: Garud, Messer et al, arXiv 2014

H: Messer et al, unpublished

Pairwise haplotype homozygosity scan



Average pairwise homozygosity tract length at genomic position p:

$$H(p) = c \times \Sigma_{i<j} h_{ij}(p)$$

Known drug resistance loci in *P. falciparum*

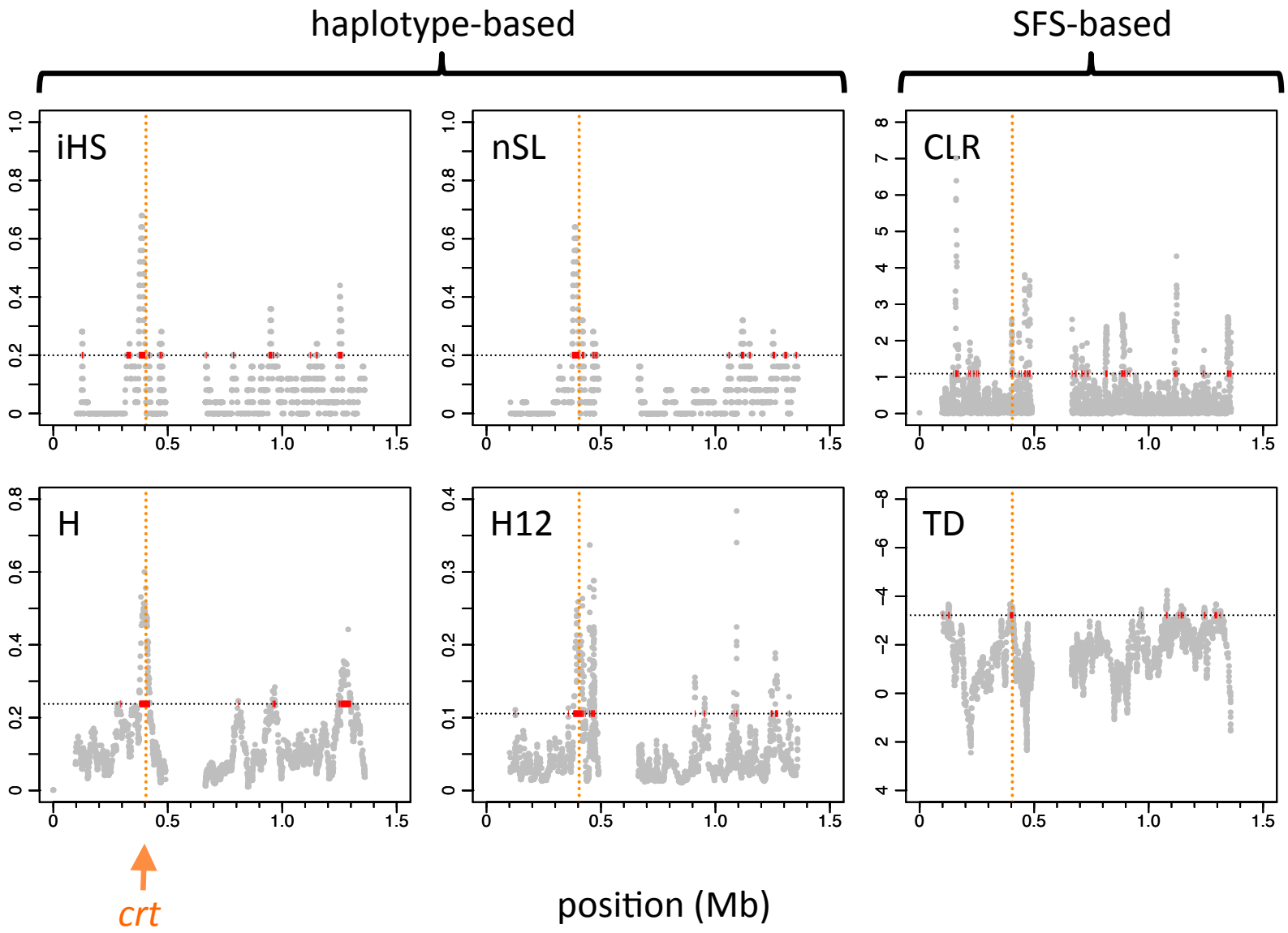
***dhfr*:**	chr 4, sequential point mutations	antifolate drugs
***dhps*:**	chr 8, sequential point mutations	
***gch1*:**	chr 12, CNV (recurrent origins)	
***crt*:**	chr 7, point mutation (K76T)	chloroquine
***mdr1*:**	chr 5, CNV (recurrent origins)	

Anderson et al, *Pharmacogenomics* (2011); Naidoo and Roper, *Trends in Parasitology* (2013)

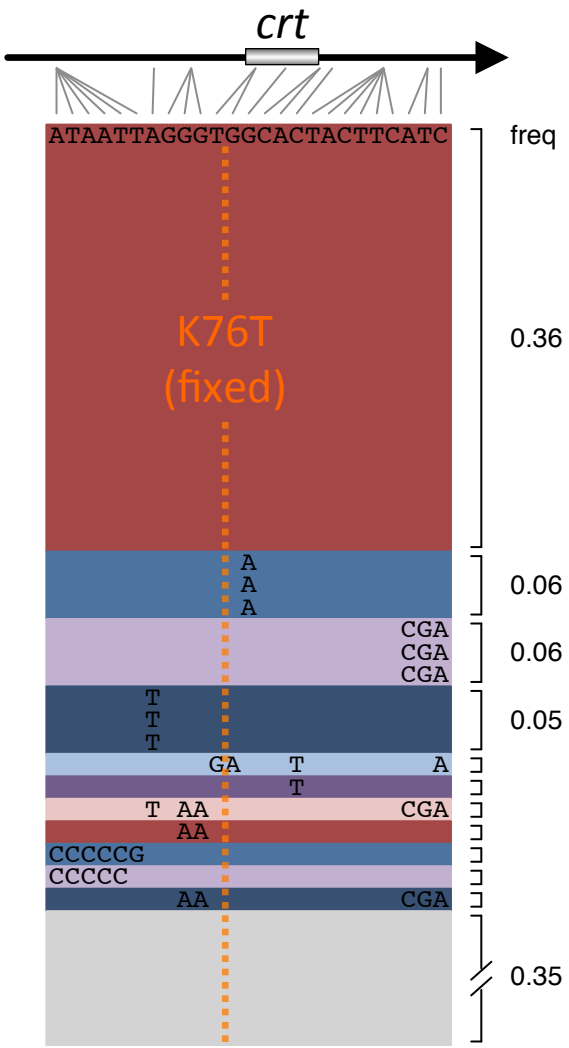
Data set

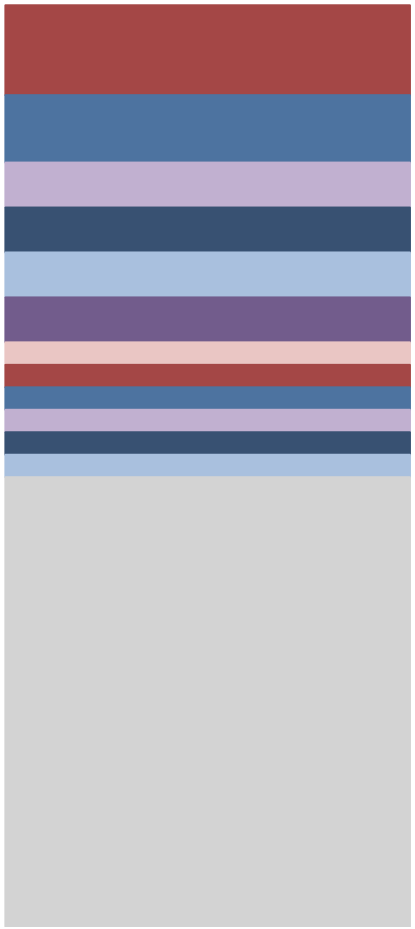
- Genome sequences of 102 *P. falciparum* parasites
- Collected at Thai-Burmese border (2003 – 2006)
- Obtained from infected patients prior to treatment
- Haploid parasite stage sequenced (no phasing)
- High-quality SNPs with > 5% MAF (32,600 total)

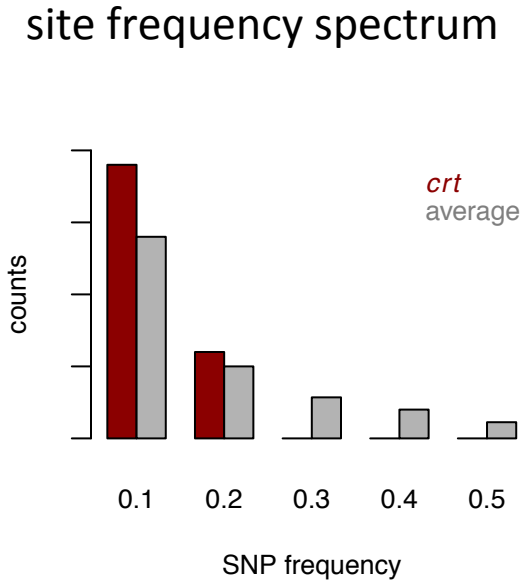
Chromosome 7 adaptive landscape



Signatures of adaptation at *crt*







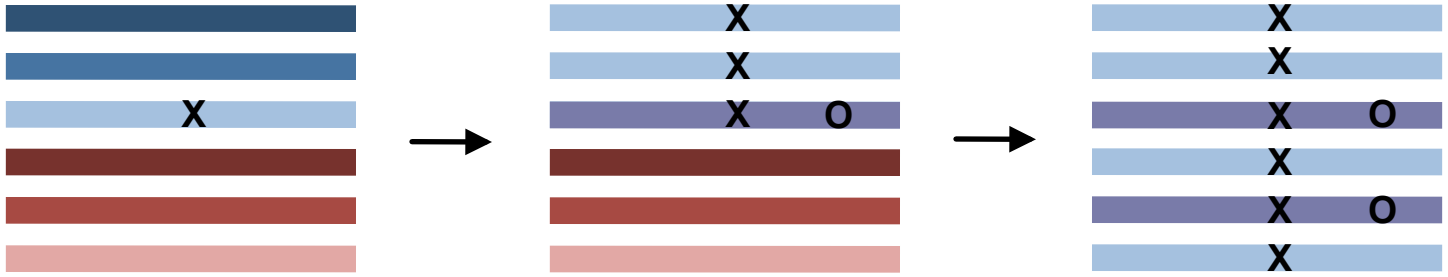
single frequent haplotype

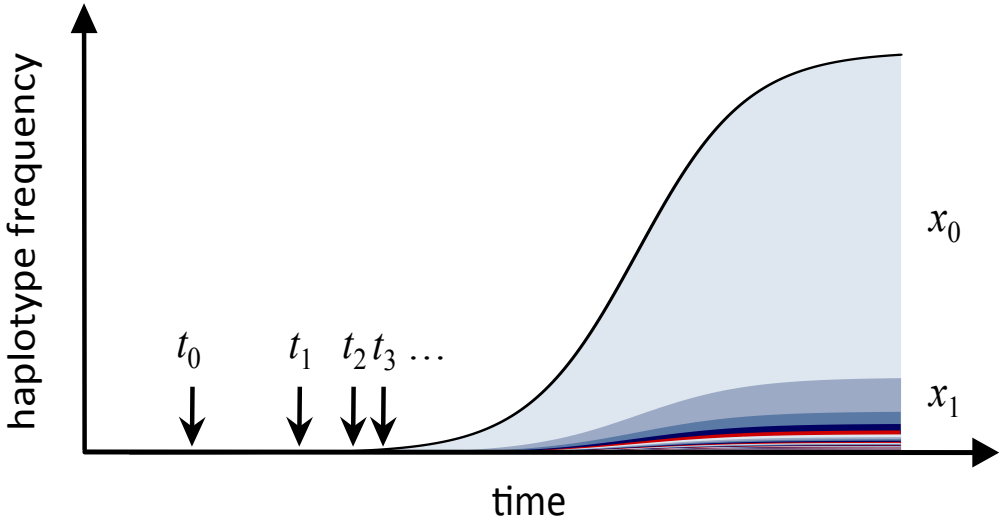
low frequency skewed SFS

↓

hard sweep

Haplotype diversity in hard sweeps

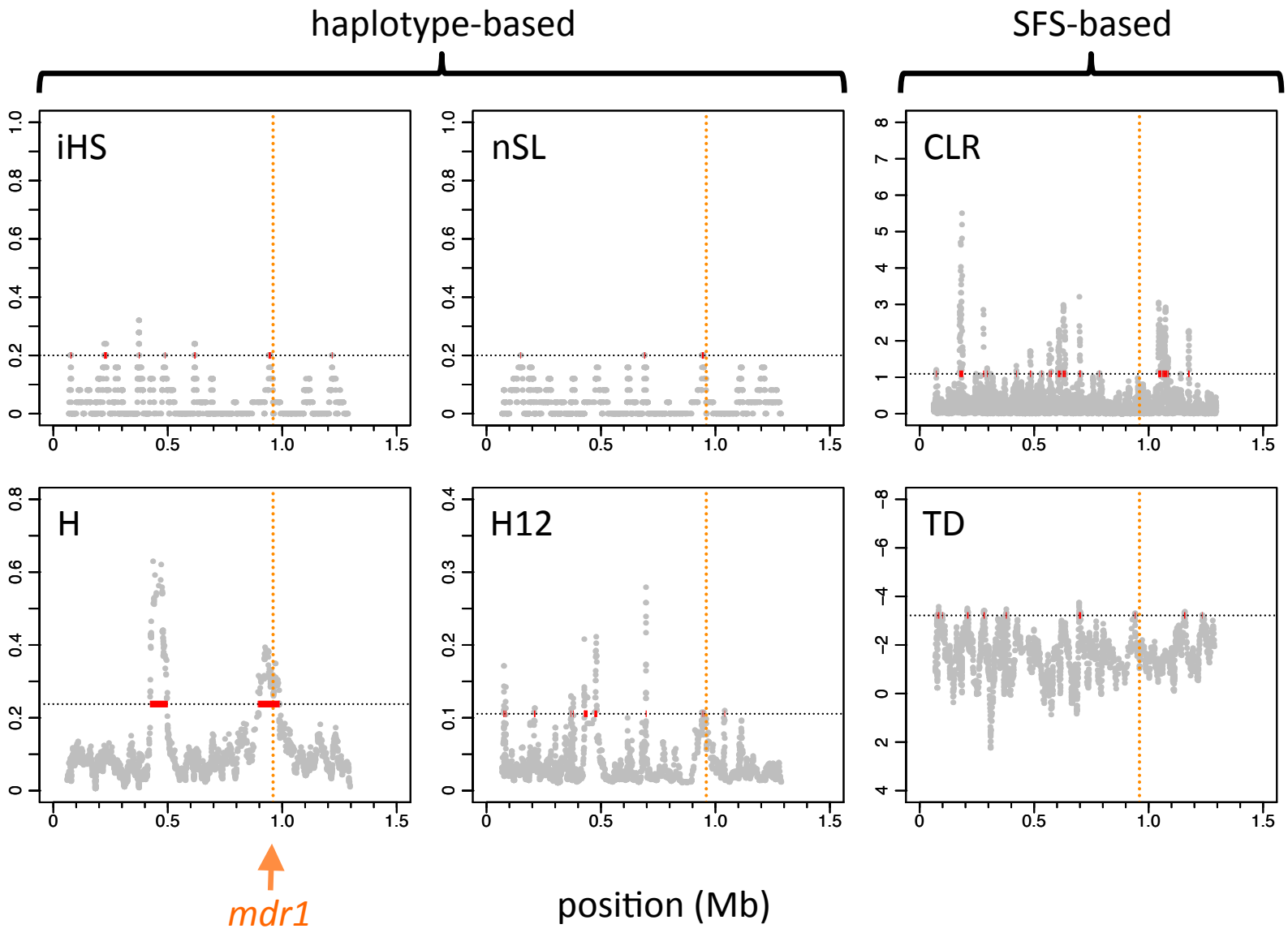




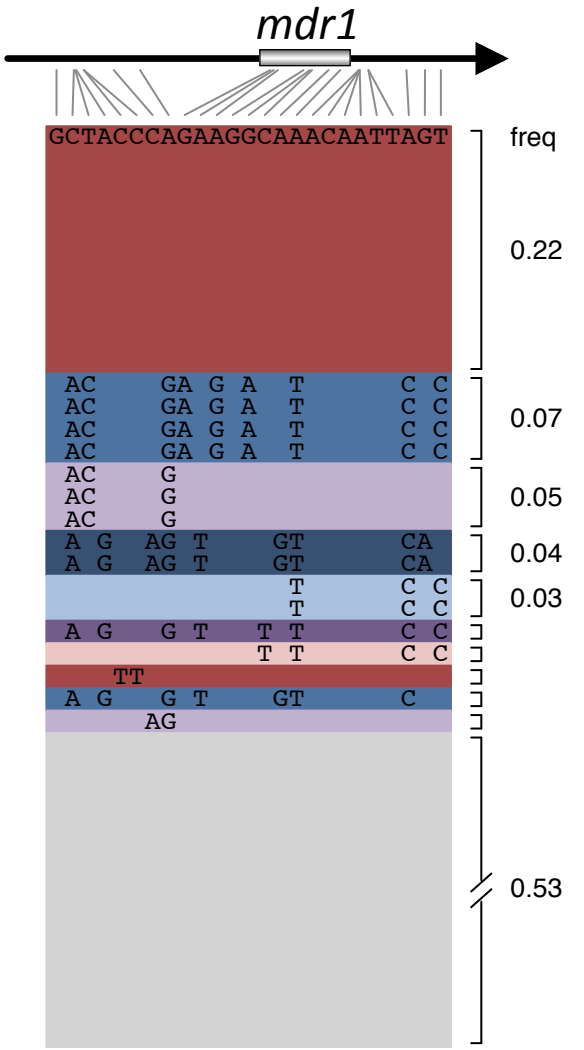
$$\mathrm{E}\left(\frac{x_i}{x_0}\right) = \frac{(\mu+\rho)L}{is}$$

Messer and Neher, *Genetics* (2012)

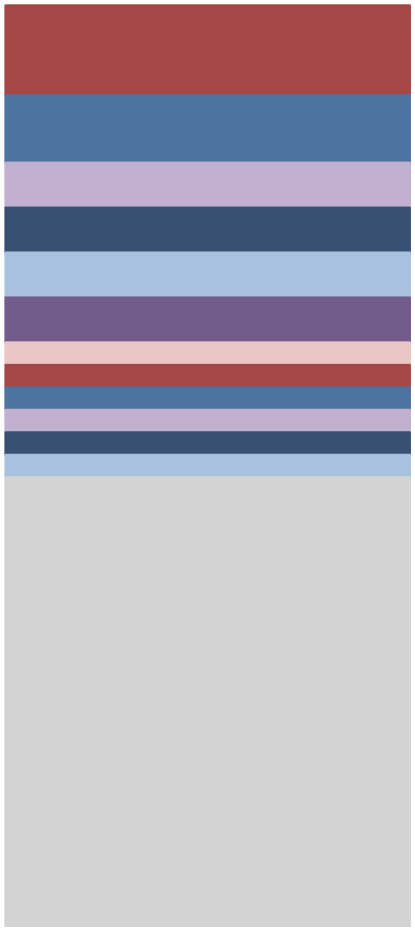
Chromosome 5 adaptive landscape



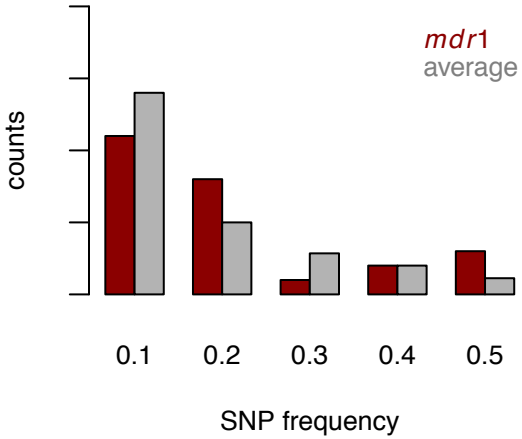
Signatures of adaptation at *mdr1*



genomic average



site frequency spectrum



several distinct haplotypes

high frequency skewed SFS

↓

soft sweep

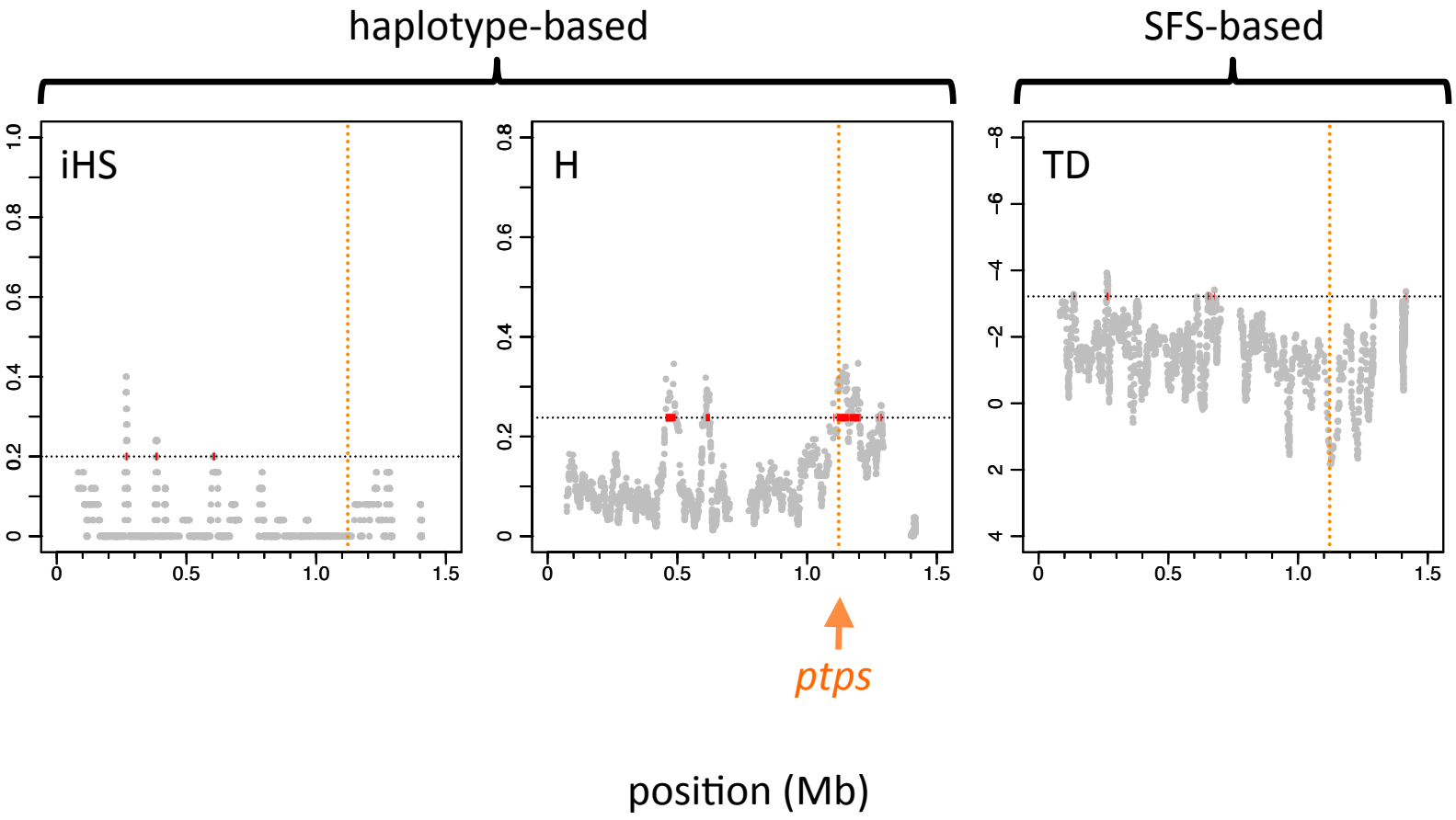
Scan performances at positive control loci

			hard	?	soft		
			crt	*dhfr*	*mdr1*	*dhps*	*gch1*
haplotype-based		H	-	0	21	0	0	0
		iHS	11 SNP	11	27	8	9	0
			25 SNP	5	26	9	1	18
			51 SNP	0	31	>50	0	24
		nSL	11 SNP	1	26	8	1	0
			25 SNP	1	26	10	1	18
			51 SNP	0	>50	>50	0	0
		H12	11 SNP	3	31	>50	4	>50
			25 SNP	0	26	9	0	27
			51 SNP	0	42	0	0	3
SFS-based		CLR	1000 bins	2	47	>50	>50	22
			5000 bins	1	40	>50	45	20
		TD	11 SNP	0	31	29	>50	>50
			25 SNP	0	>50	16	>50	>50
			51 SNP	0	>50	>50	>50	>50

d=0
0<d≤25
25<d≤50
d>50

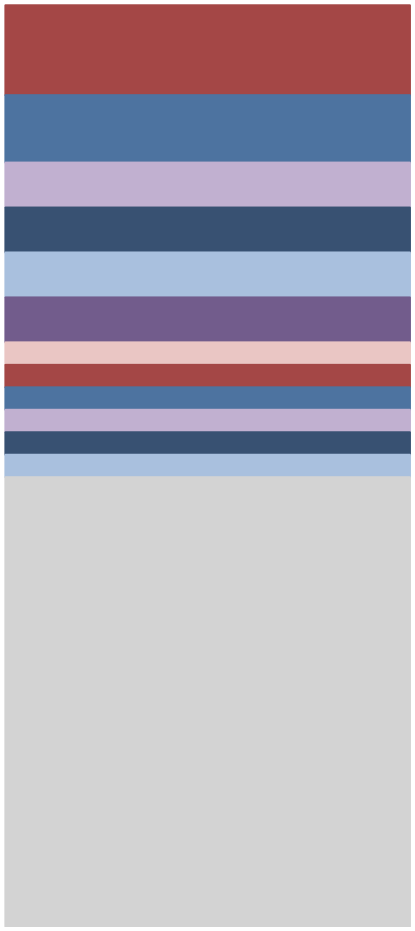
Numbers show distance (kb) to nearest detected locus (0.95 quantile criterion)

Chromosome 6 adaptive landscape

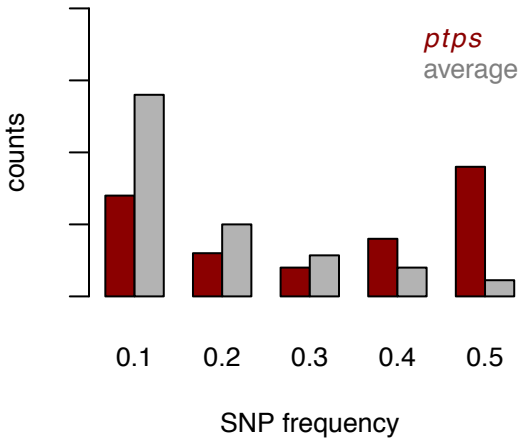


Signatures of adaptation at *ptps*

genomic average



site frequency spectrum



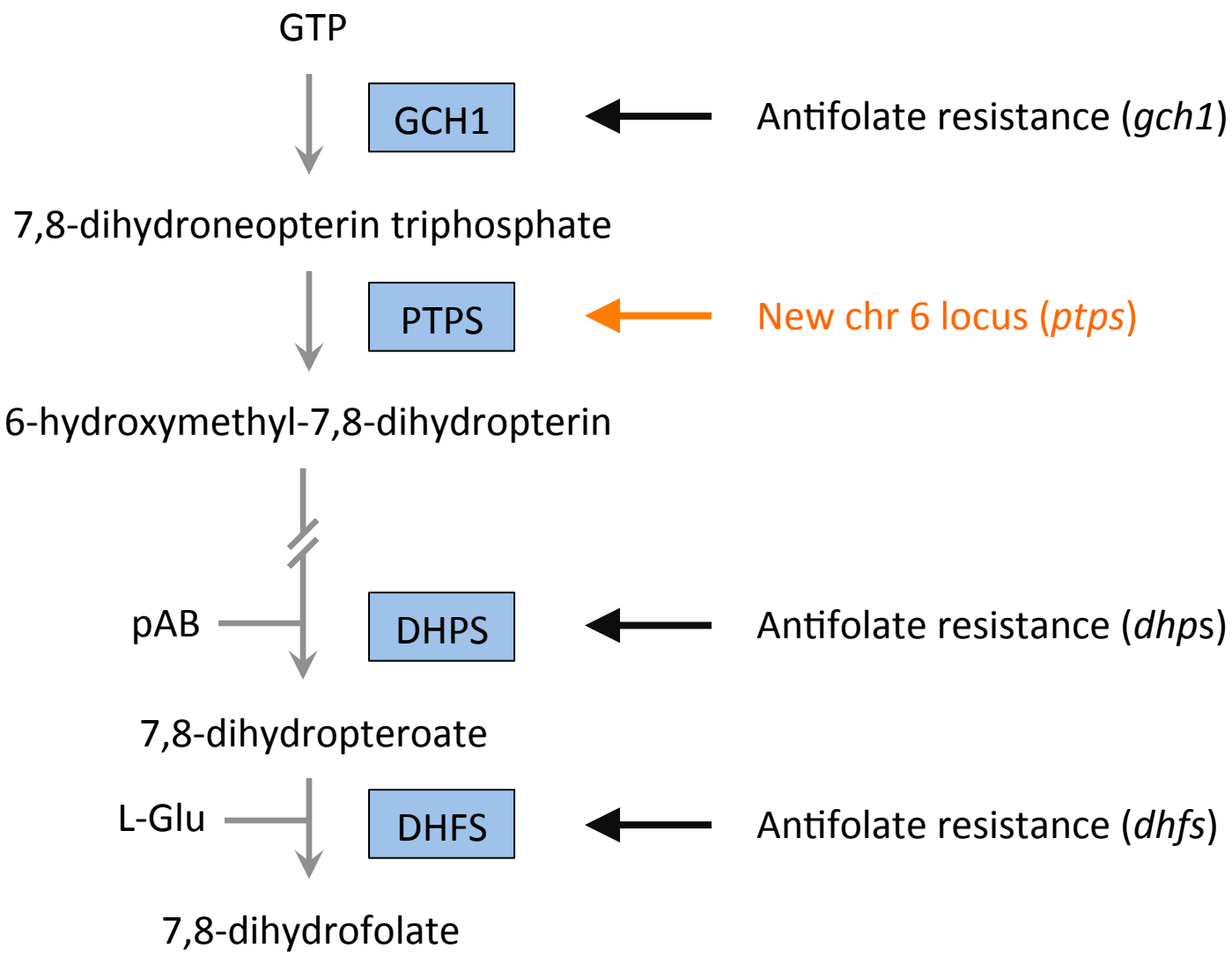
several distinct haplotypes

high frequency skewed SFS

↓

soft sweep

Folate pathway and drug resistance



Summary

- Evolution of drug resistance in *P. falciparum* produced detectable selective sweeps
- Haplotype-based statistics can in principle detect all presently known resistance loci
- SFS-based methods fail at detecting most loci
- Discrepancy likely reflects the fact that drug resistance in Malaria commonly produced soft sweeps

Acknowledgments

Stanford:

Dmitri Petrov
Ben Wilson
Nandita Garud
Pleuni Pennings
Erkan Buzbas

Texas Biomed:

Ian Cheeseman
Shalini Nair
Tim Anderson

Sanger Institute:

Dominic Kwiatkowski
Magnus Manske
Bronwyn MacInnis
Gareth Turner

SMRU Thailand:

Francois Nosten

MPI Tuebingen:

Richard Neher