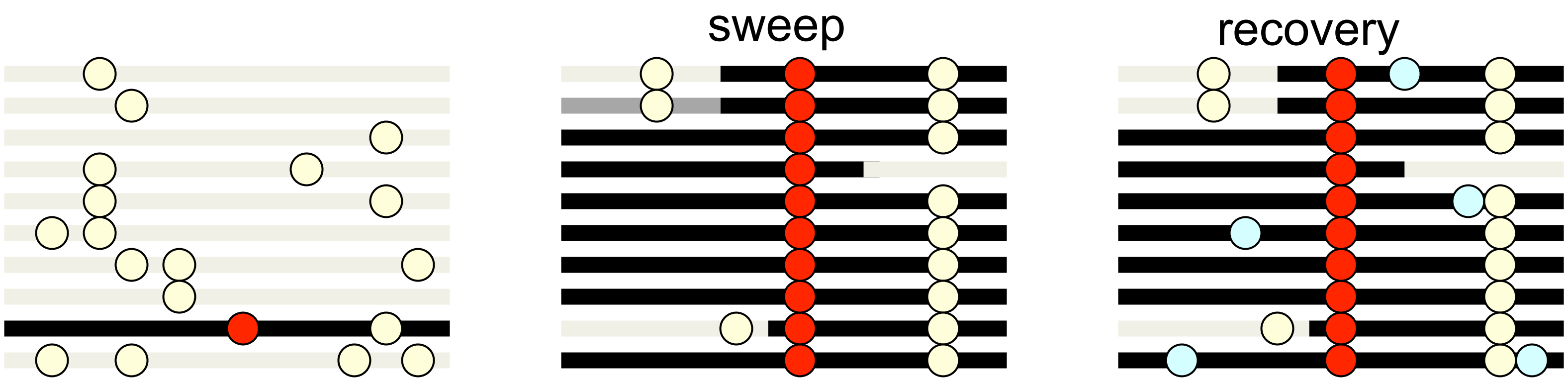
The effect of partial sweeps on diversity

A partial draft of hitchhiking

Graham Coop and Peter Ralph
UC Davis

The effect of selective sweeps on linked neutral variants

Maynard Smith and Haigh, Kaplan et al, Braverman et al etc

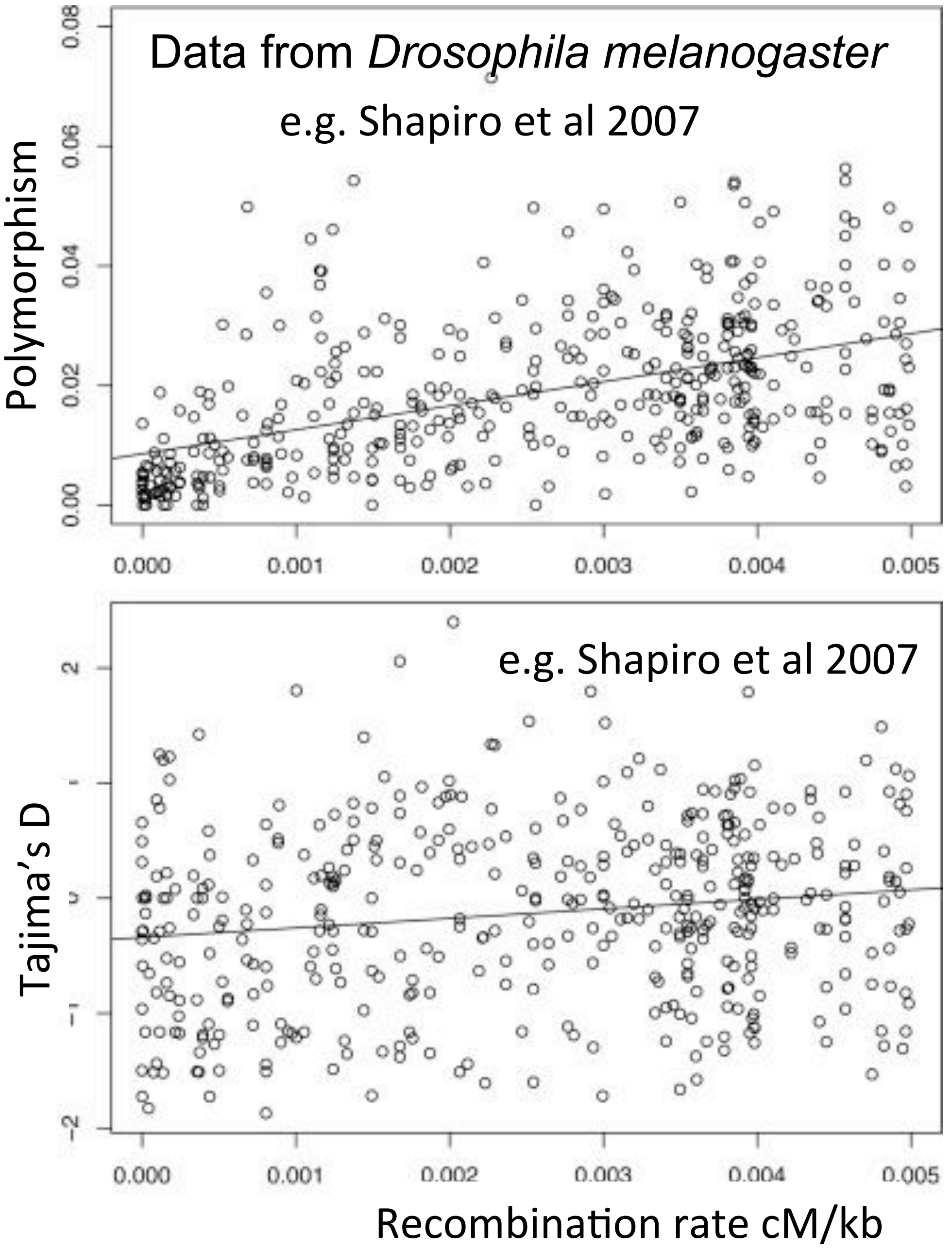


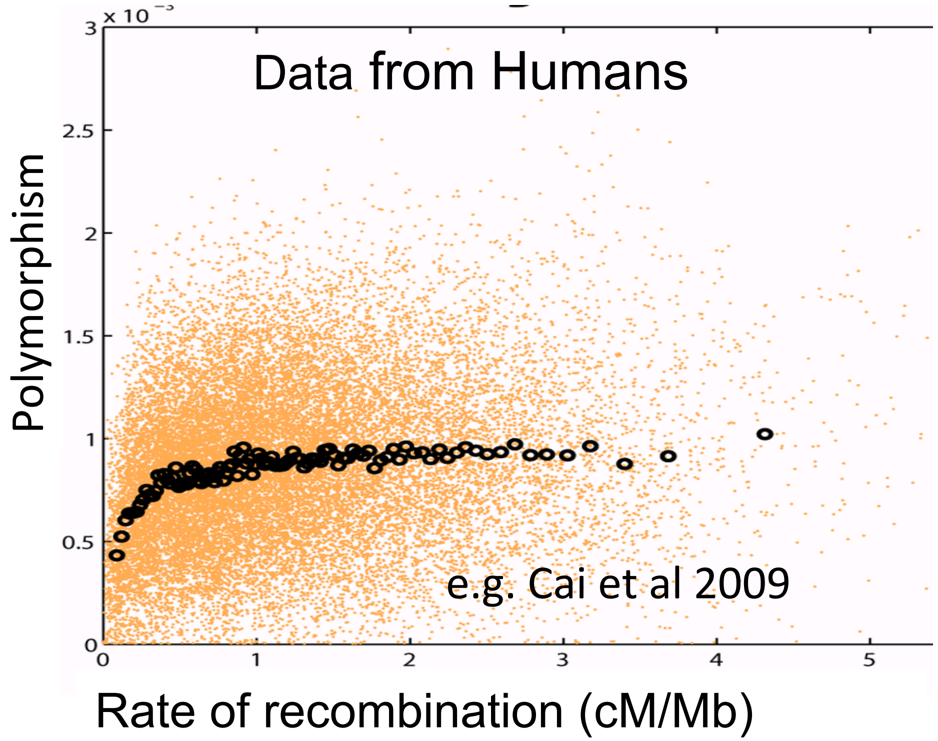
Reduced diversity
high frequency derived
alleles

New mutations
lead to a skew towards
rare alleles

Background selection can also lead to a reduction in diversity,
but lead to only a weak skew towards rare alleles

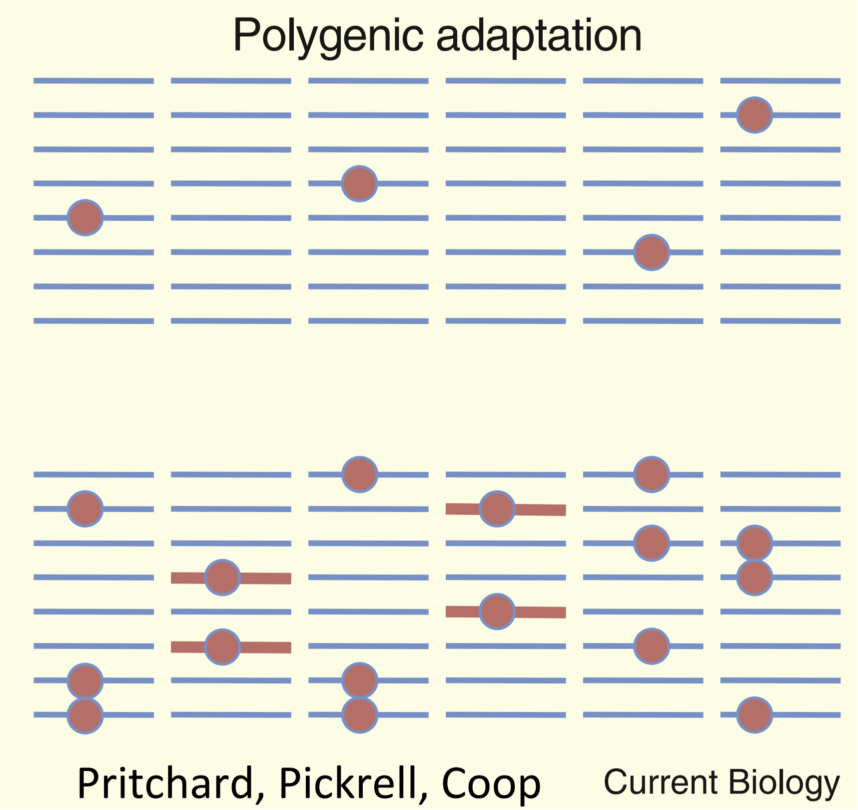
Evidence for linked selection





Only a very weak skew towards singletons in low recombination regions in humans.

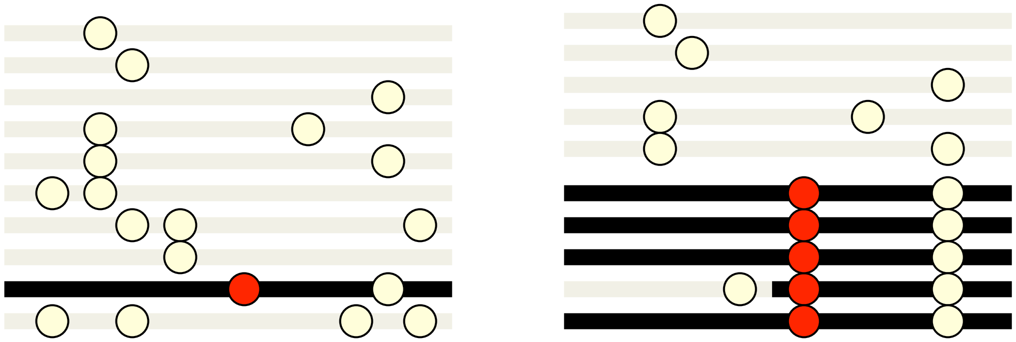
What if most newly arisen selected alleles do not sweep to rapidly fixation?
E.g. due to changing environment or genomic background
(Due to parallel mutation, other variation etc)



Pritchard, Pickrell, Coop Current Biology

E.g. Chevin and Hospital

What if most newly arisen selected alleles do not sweep to rapidly fixation?
E.g. due to changing environment or genomic background
(Due to parallel mutation, other variation etc)



The coalescent process with a partial sweep

A model of recurrent partial sweeps

Model where these sweeps occur along the genome

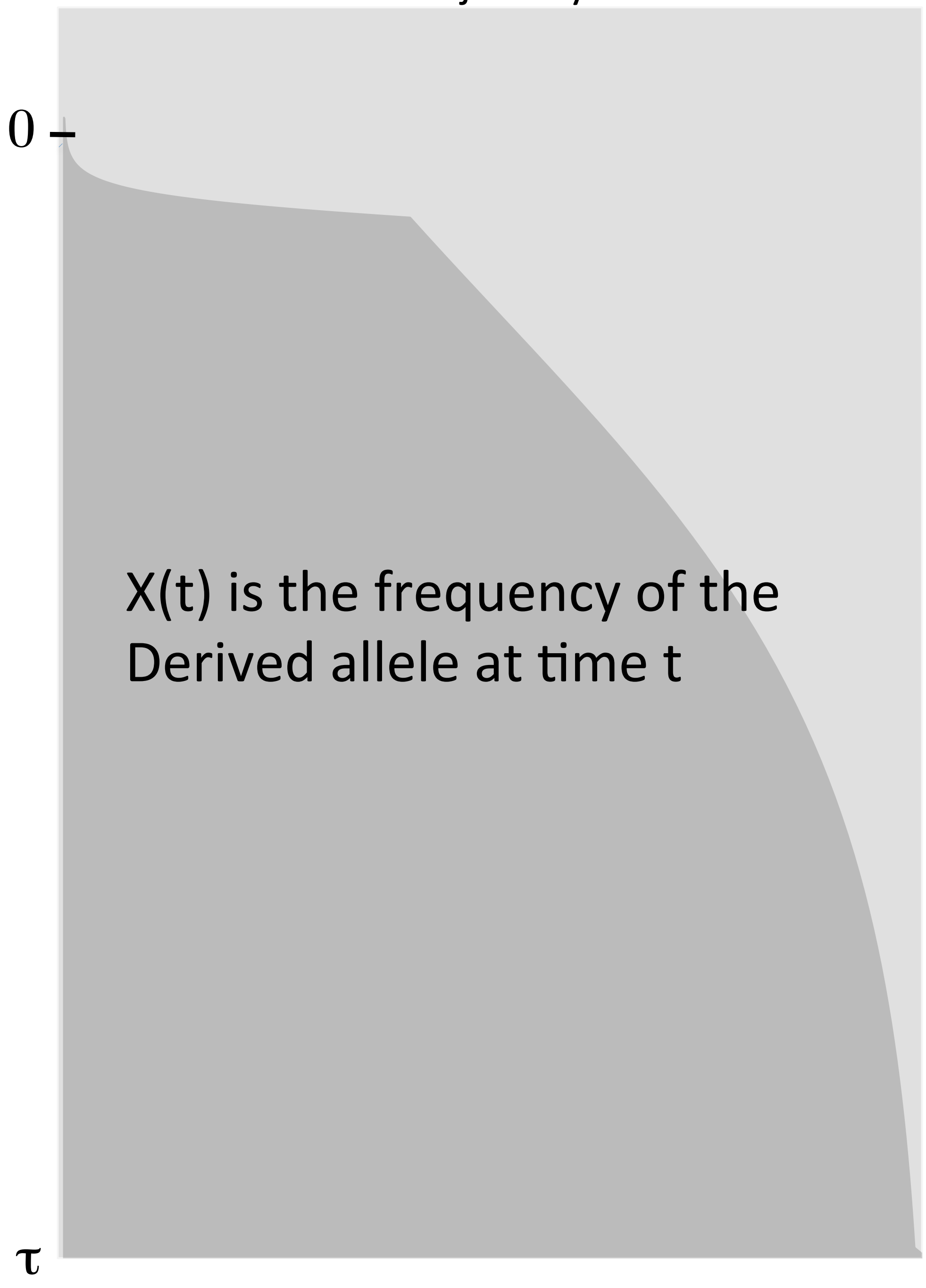
-Effect on diversity

-Effect on frequency spectrum

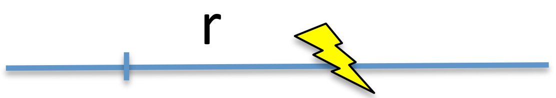
The derived allele arose τ Generations ago

Conditions on trajectory:
Selected allele initially
quickly increases
In frequency.

B. Coalescent with trajectory

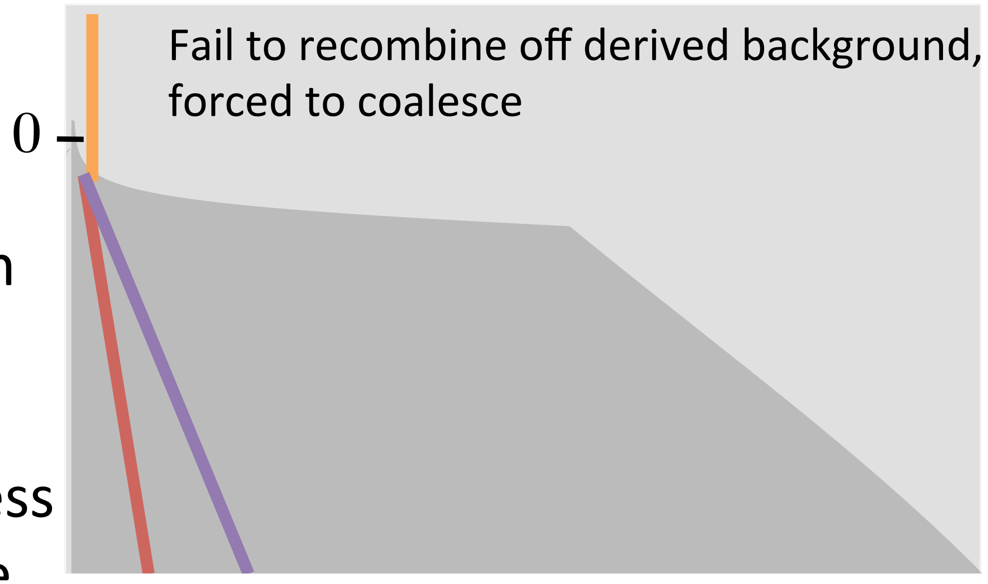


X(t) is the frequency of the Derived allele at time t

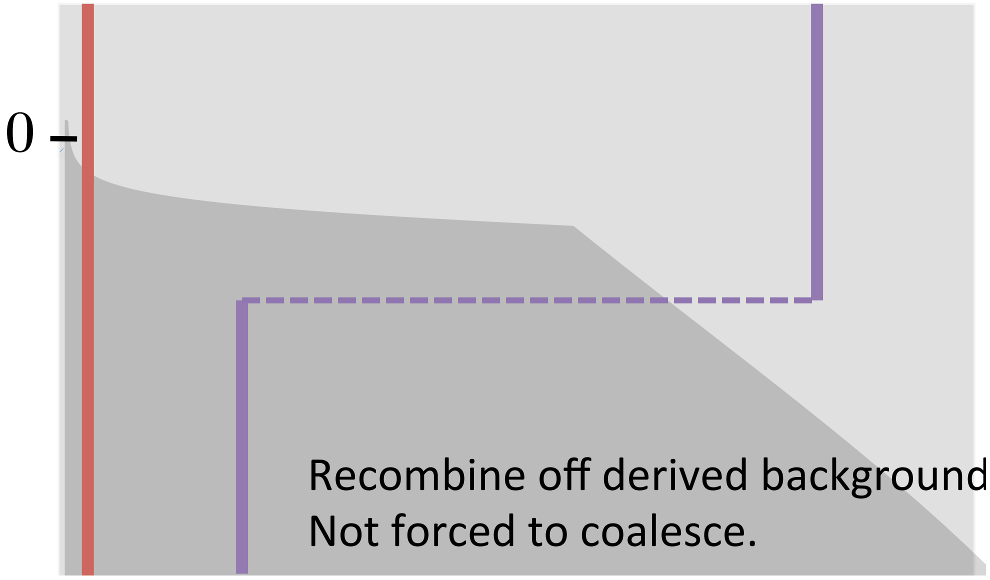


neutral site
a genetic distance r away from
the selected locus

Consider the coalescent process
of a pair of lineages at this site

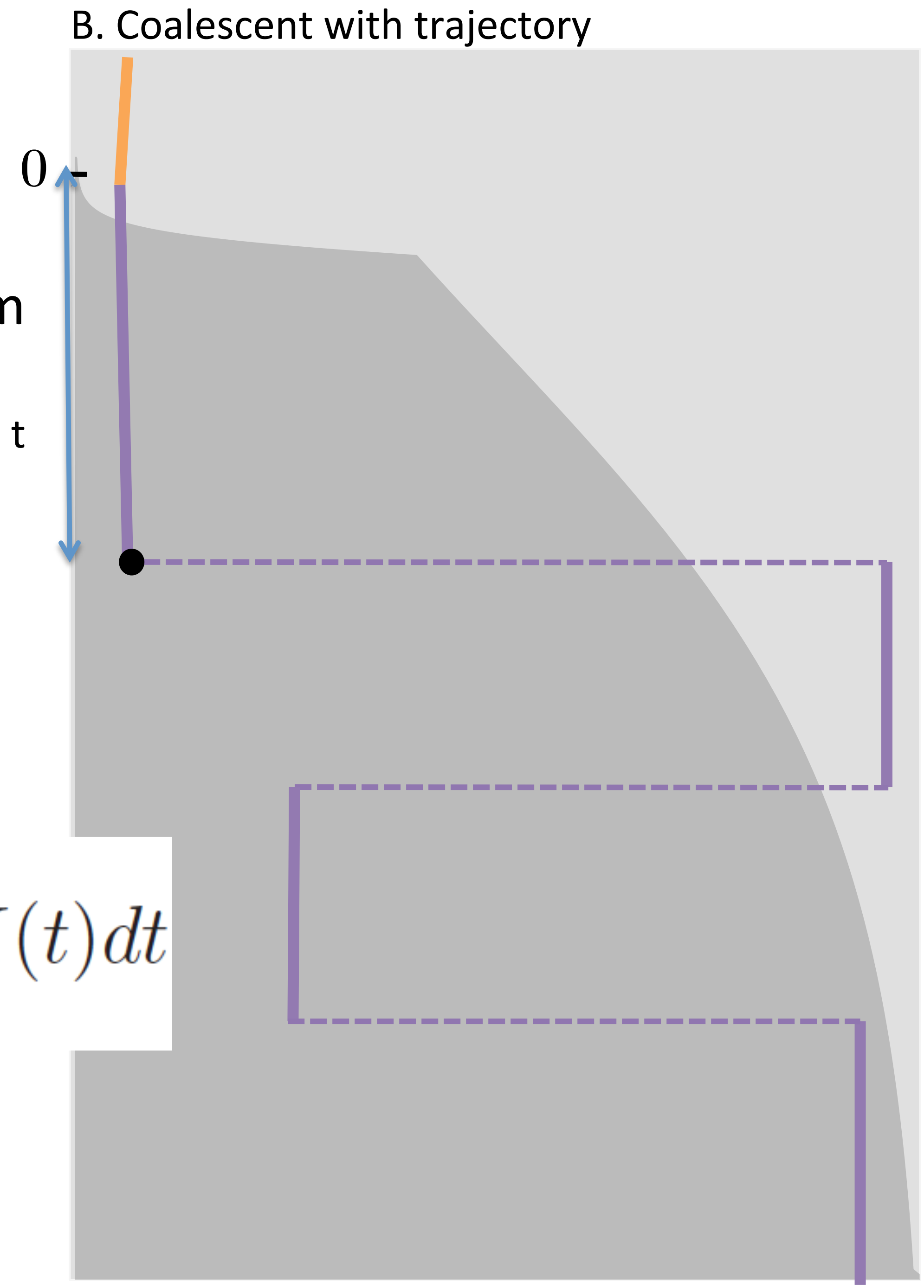


Fail to recombine off derived background,
forced to coalesce



Recombine off derived background
Not forced to coalesce.

Imagine a neutral site
a genetic distance r away from
the selected locus

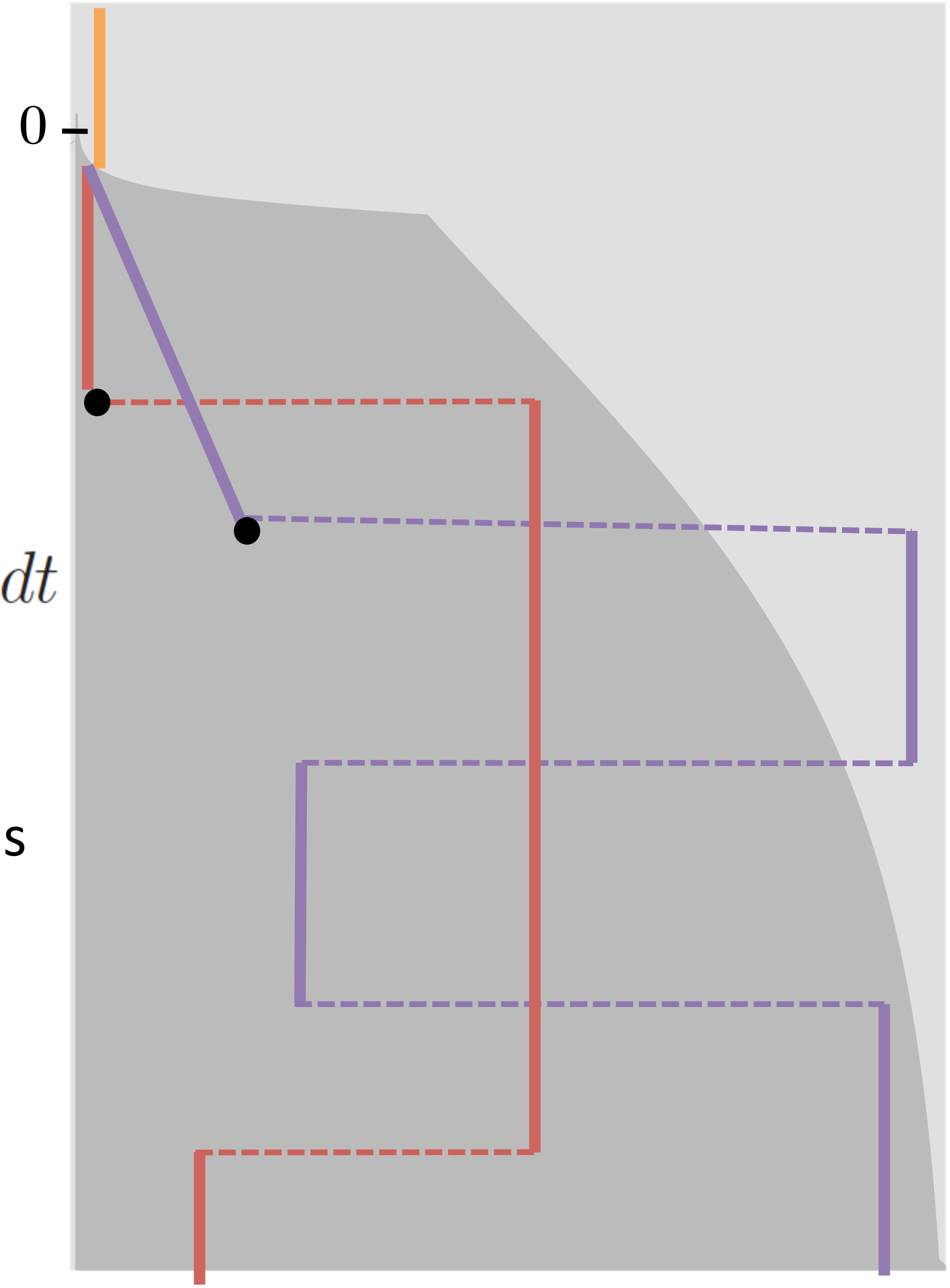


Probability that the lineage is of the derived type at time 0 =

$$q(r, X) = r\int_0^{\tau} e^{-rt} X(t)dt$$

$r\,\tau >>1$

B. Coalescent with trajectory



$$q(r, X) = r\int_0^{\tau} e^{-rt}X(t)dt$$

Probability that 2 lineages
Are forced to coalesce
At time 0 by sweep = q^2

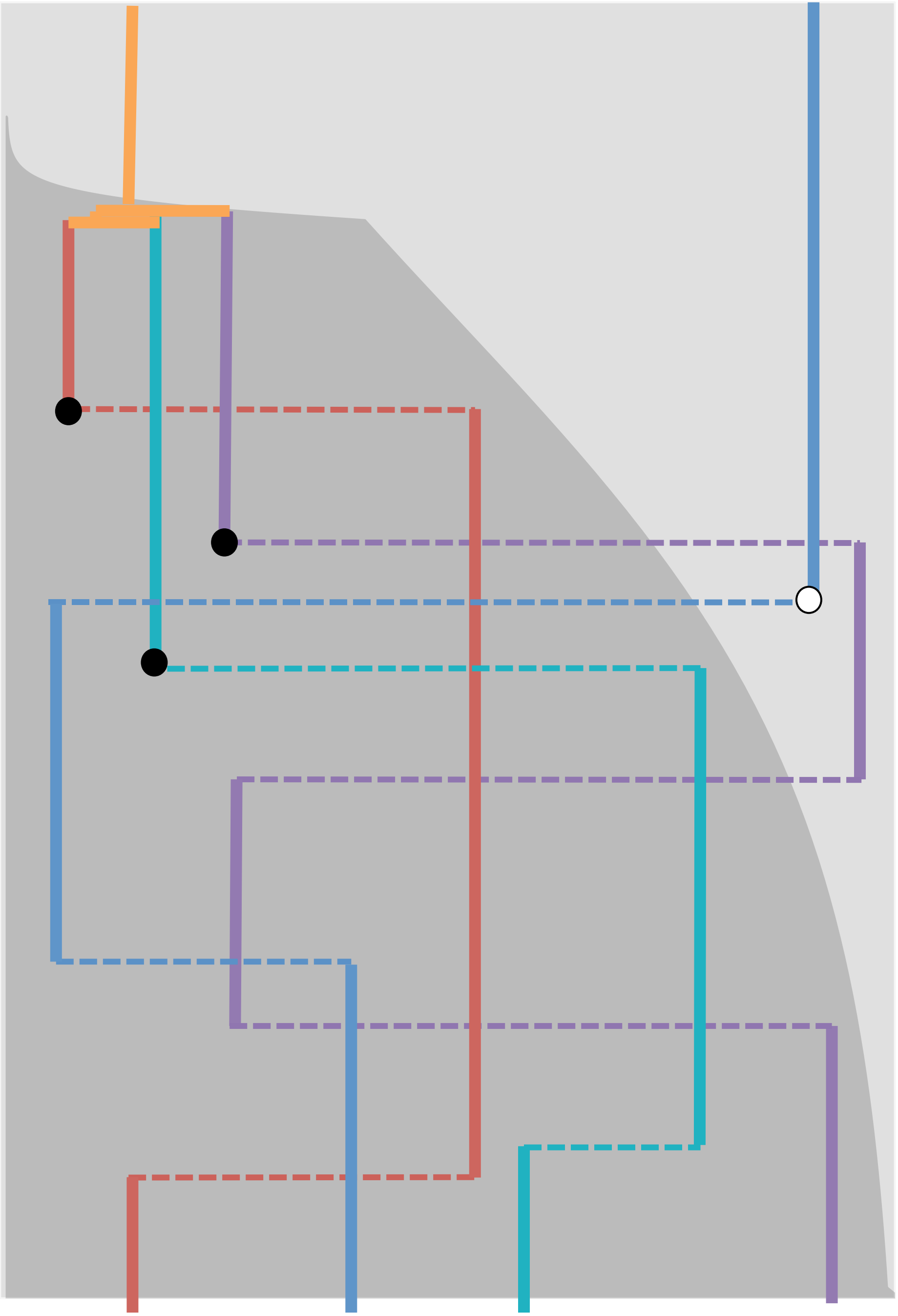
$$q(r, X) = r \int_0^{\tau} e^{-rt} X(t) dt$$

Probability that out of k lineages i are force to coalesce is binomial:

$$\binom{k}{i} q^i (1-q)^{k-i},$$

$$\text{for} \quad 2 \leq i \leq k,$$

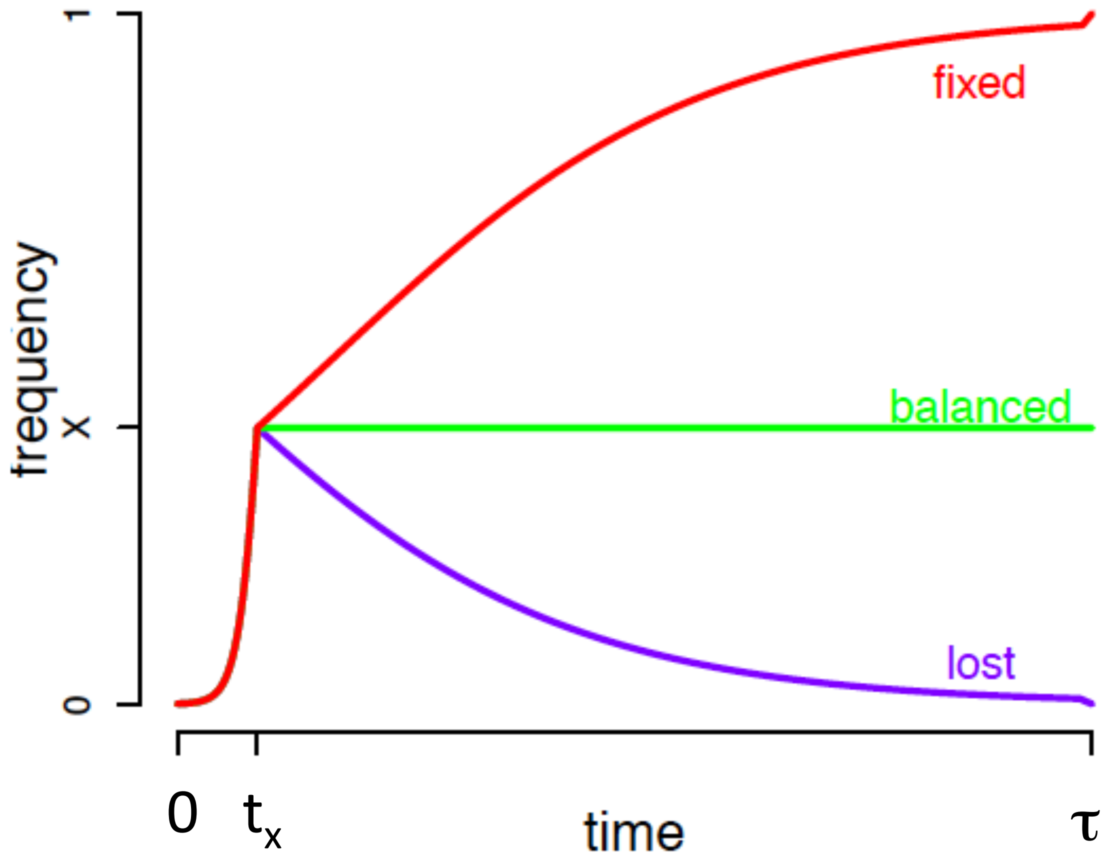
B. Coalescent with trajectory



Simple trajectories

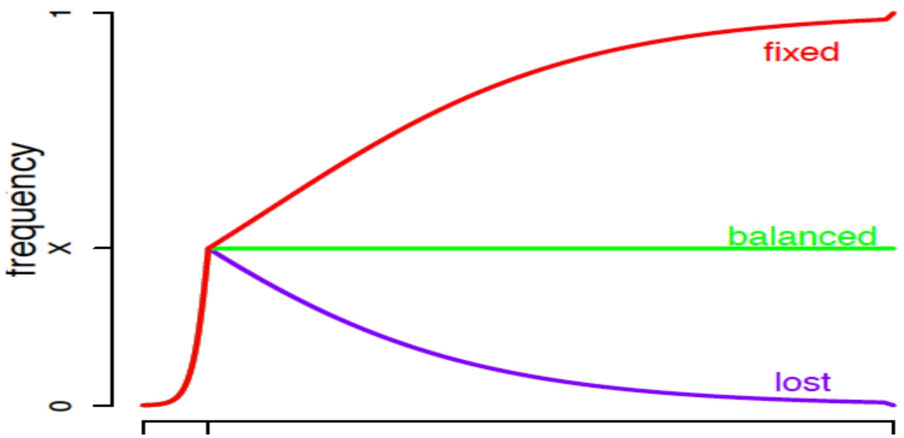
Selected allele moves quickly from 1/2N to x in time t_x

Then stays at x, or goes to fixation, or loss on a slower time-scale (e.g. with selection coefficient s_2, $-s_2$, or 0 respectively)



$$q \approx x e^{-r t_x}$$

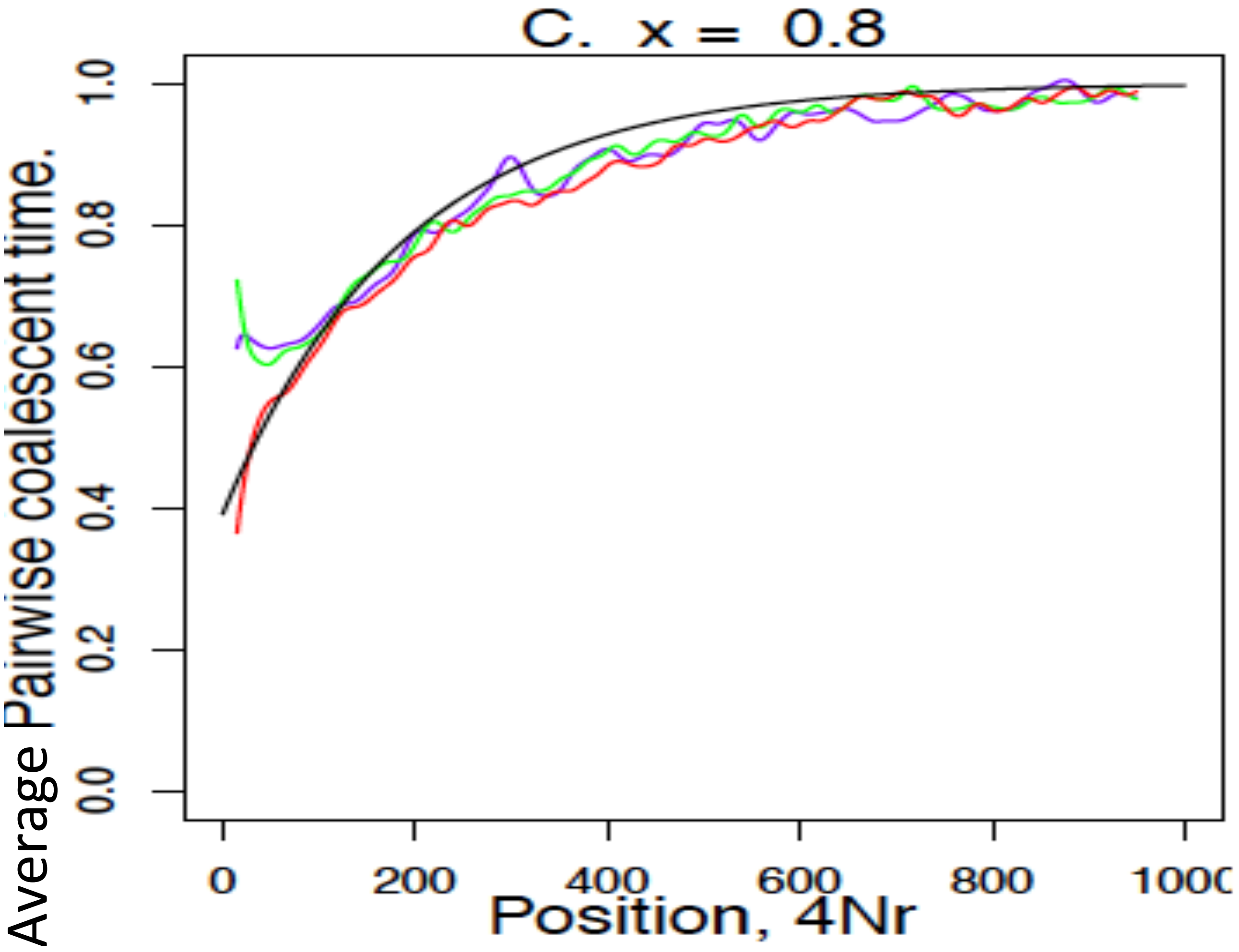
Also holds for other Trajectories when $r >> s_2$



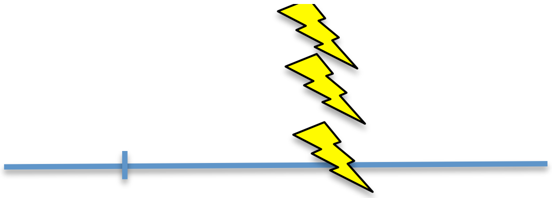
$$2N(1 - q_x^2 e^{\tau/(2N)})$$

$t_x /2N = 0.0053$

$\tau/2N = 0.05$



Recurrent sweep process



- Sweeps happen at rate ν
- Rate of sweep coalescence $\nu\, q^2$
- Neutral rate of coalescence: 1/(2N)
- Total rate of coalescence $\nu q^2 + 1/(2N)$

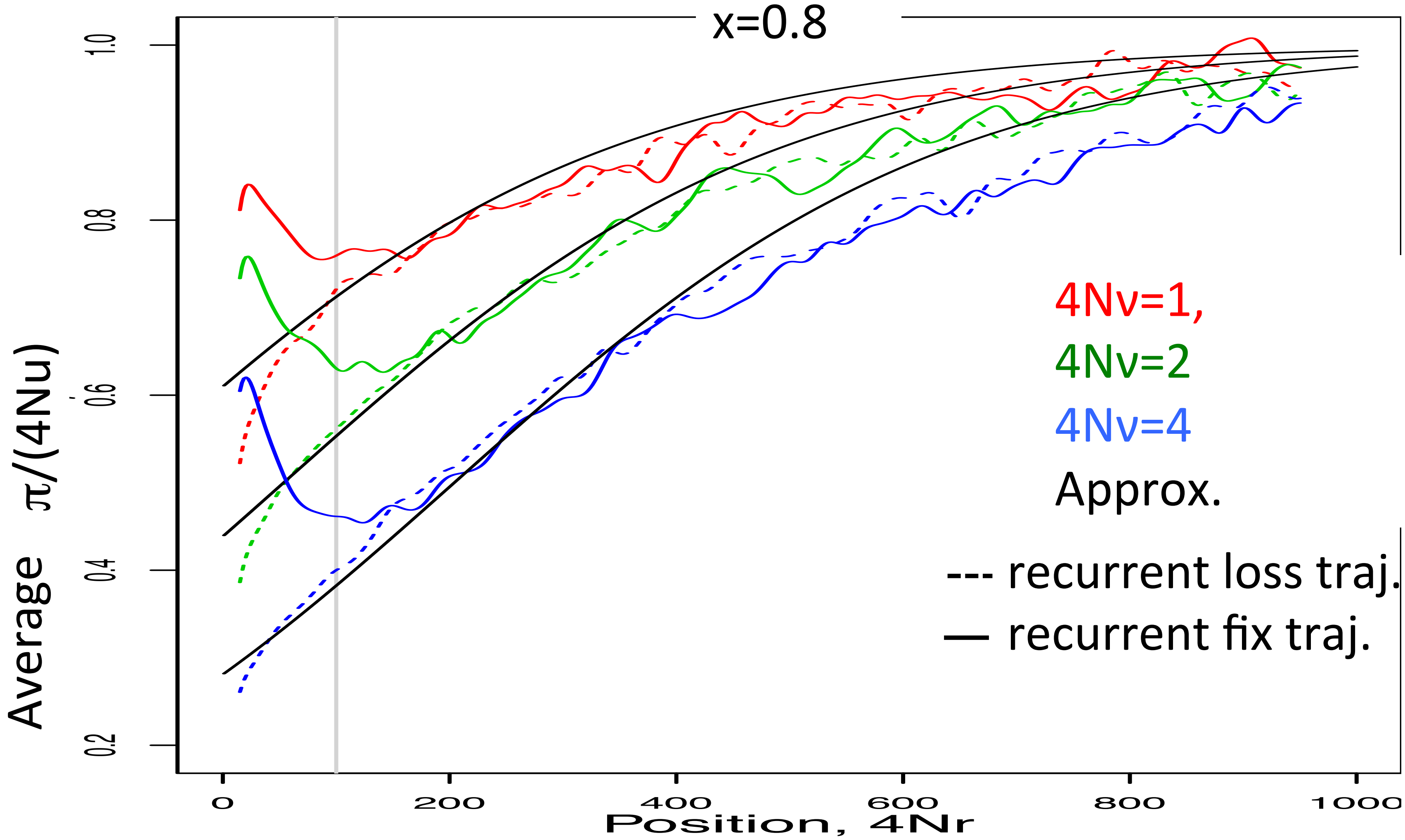
$$\mathbb{E}(T) = \frac{2N}{1 + 2N\nu q^2} \qquad \mathbb{E}(\pi) = \frac{4Nu}{1 + 2N\nu q^2}$$

u=mutation rate at neutral locus.

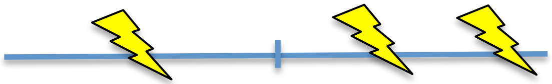
$$\mathbb{E}(\pi) = \frac{4Nu}{1 + 2N\nu q^2}$$

u= mutation rate

- For our simple approximation $q \approx xe^{-rt_x}$



Homogeneous sweeps



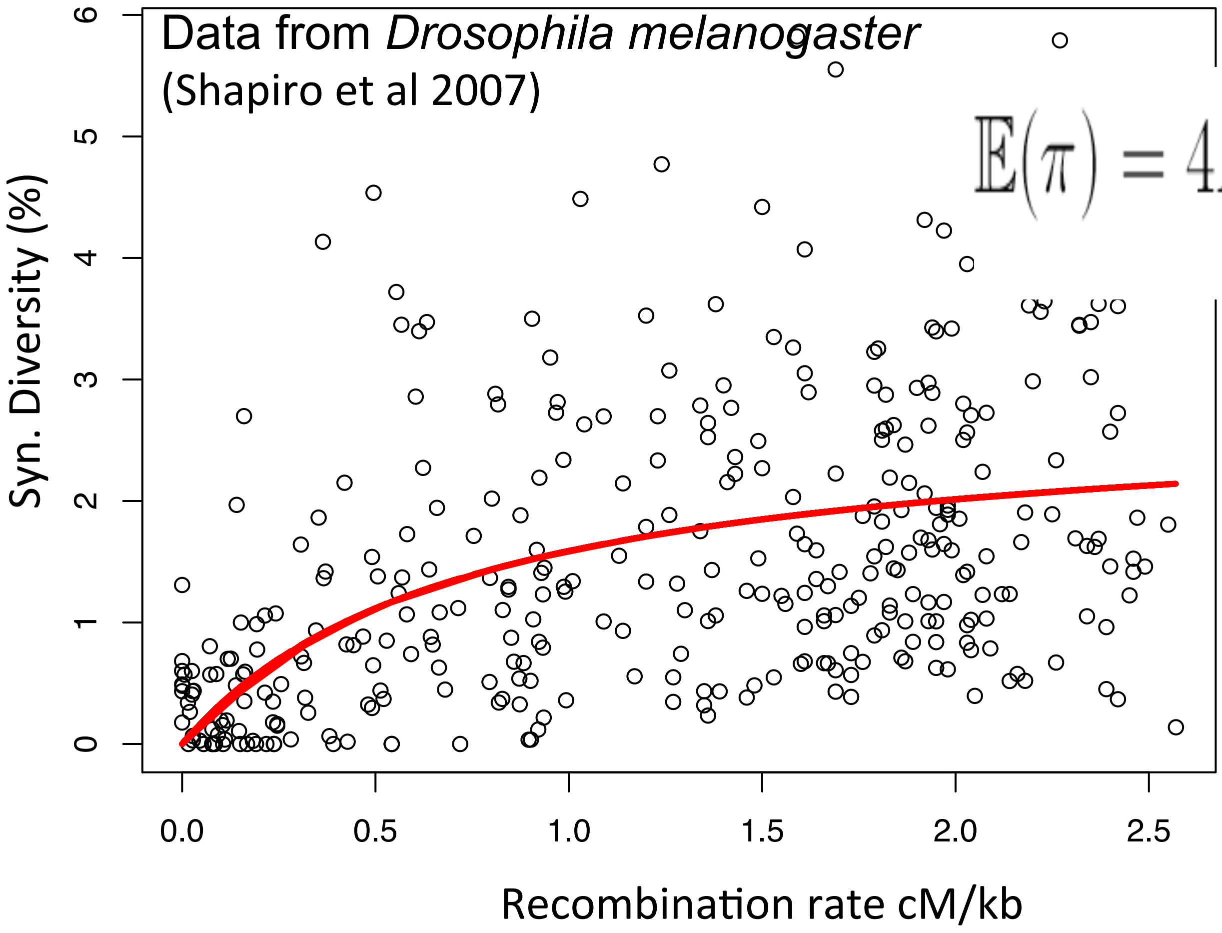
Infinitely long chromosome recombining at rate r_{BP}
Sweeps occur homogeneous at rate ν_{BP} per generation

$$\mathbb{E}(T) = \frac{2Nr_{BP}}{r_{BP} + 2N\nu_{BP}J}$$

Where J depends only on the form taken by the trajectory.

$$J = \int_0^\infty q(r)^2 dr$$

Under our simple partial sweep model: $J = x^2/t_x$



Data from *Drosophila melanogaster*
(Shapiro et al 2007)

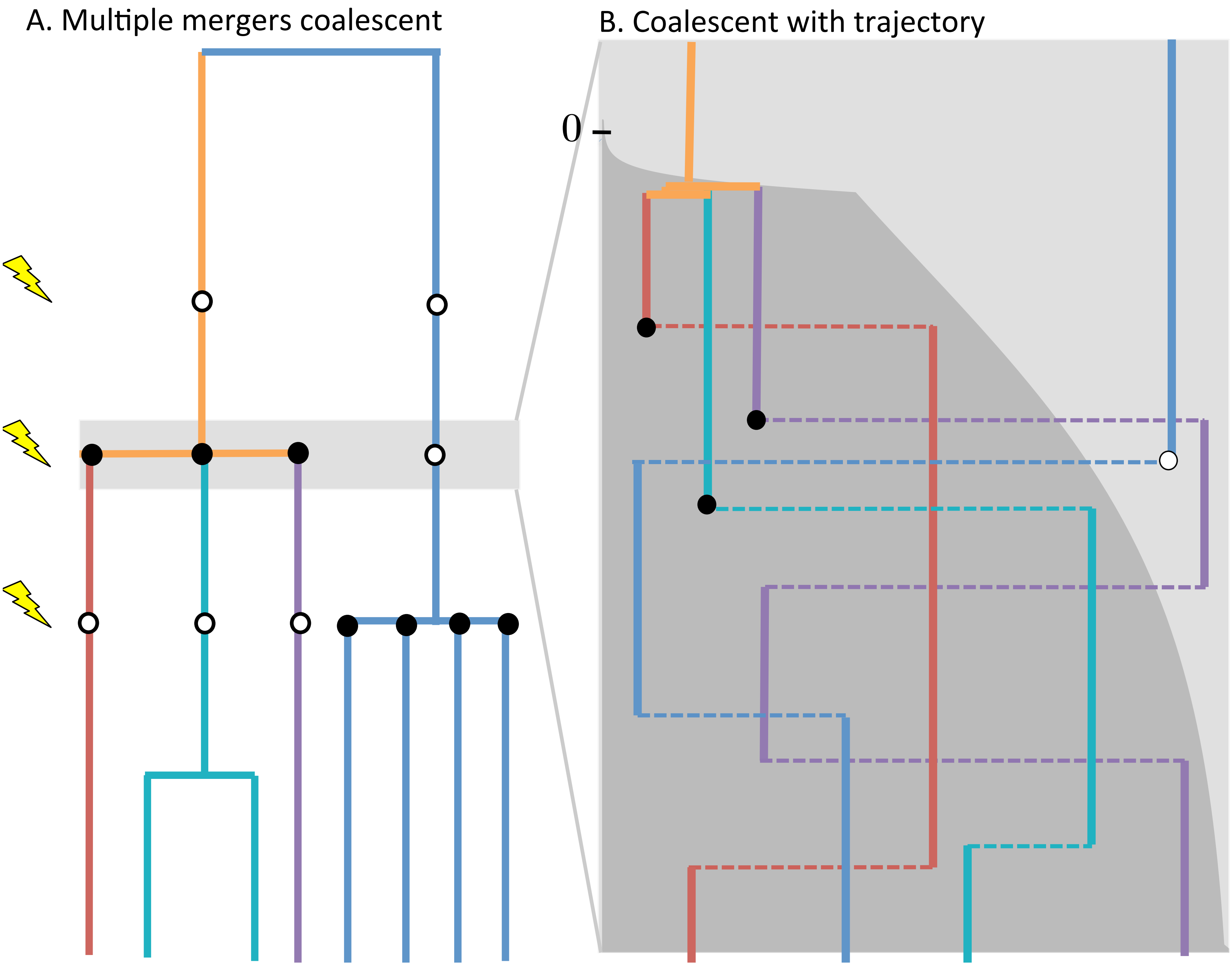
$$\mathbb{E}(\pi) = 4Nu\frac{r_{BP}}{r_{BP}+2N\nu_{BP}(x^2/t_x)}$$

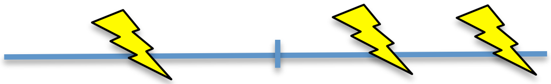
$2Nv_{BP}\ x^2/t_x=7x10^{-9}$

Assuming none of the reduction is due to Background Selection

t_x = 1000 gens (s~0.1%)
$N=10^6$
$v_{BP}\ x^2 = 3x10^{-13}$

x =	100%	20%	5%
v_{BP}=	3e-13	8e-12	1e-10


A. Multiple mergers coalescent
B. Coalescent with trajectory
0



Homogeneous sweeps at rate v_{BP} ,recombination at rate r_{BP}. Then i out of k lineages coalesce at rate:

$$= \frac{\nu_{BP}}{r_{BP}} J_{k,i}$$

$$\text{for } 2 \leq i \leq k,$$

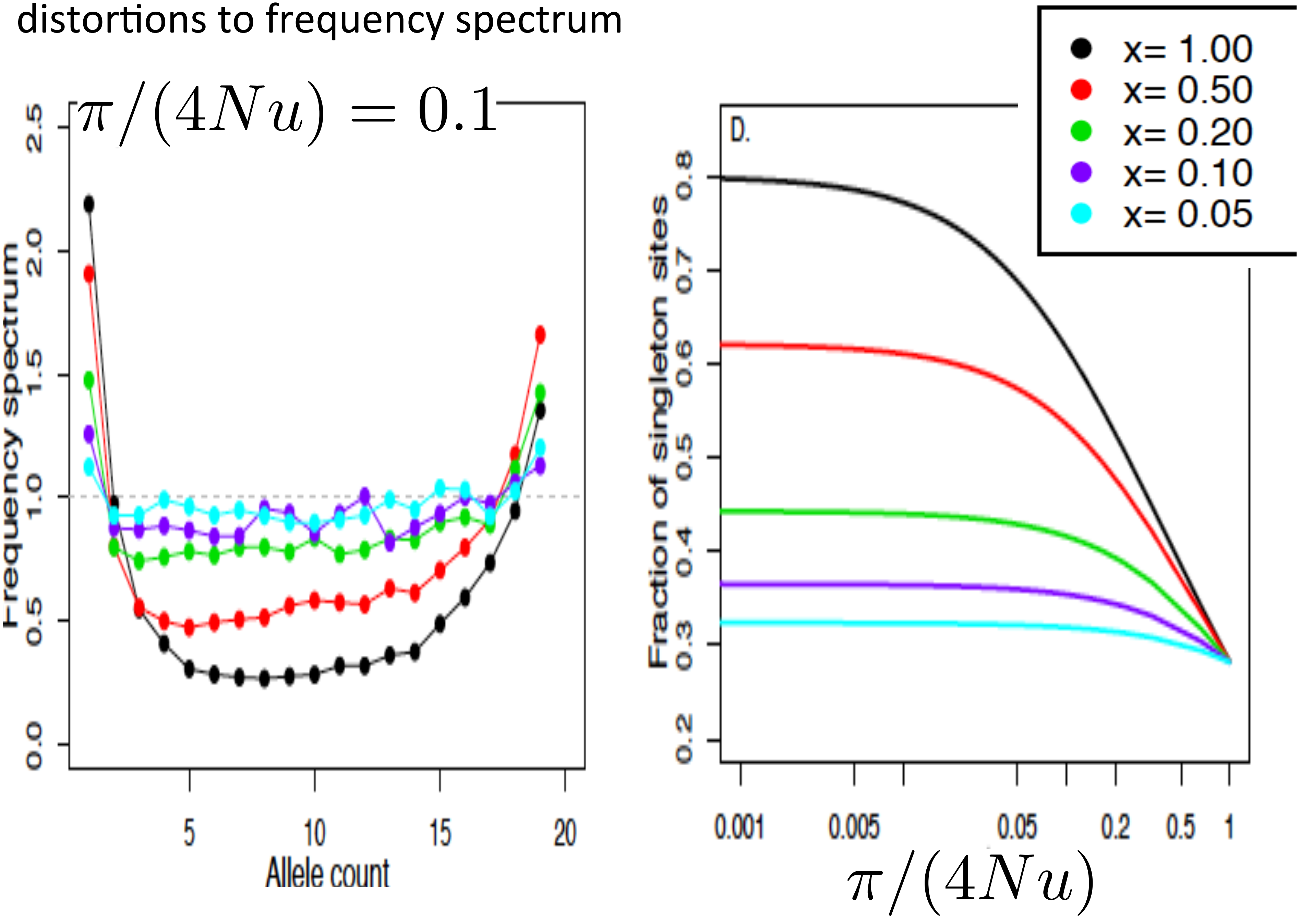
Where $J_{k,i}$ depend only on the form taken by trajectories

E.g. for our simple trajectory J is a function of x (freq. sweeps achieve)

So rate of coalescence controlled by $\frac{\nu_{BP}}{r_{BP}}$

And number of lineages forced to coalesce by x (or distribution on x).

For same reduction in diversity we can get very different distortions to frequency spectrum



Conclusions

- Partial sweeps may be responsible for reduction in diversity in low recombination regions
- If they only sweep to low frequency may result in little distortion in neutral frequency spectrum
- Relaxing our assumptions about sweeps leads to a (depressingly) much broader set of predictions.
- Idea: Rather than estimating one model why not estimate rates of different types of coalescence ($J_{k,i}$).

Thanks

Peter Ralph

Thanks to Yaniv Brandvain, Chuck Langley, Molly Przeworski, Alisa Sedghifar, and Guy Sella for helpful conversations