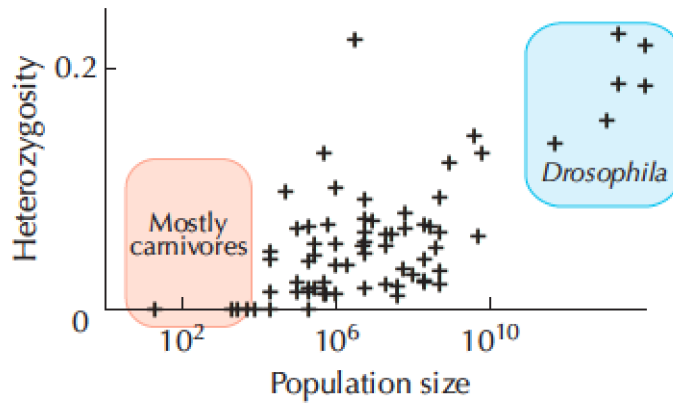

Hitch-hiking in space

Nick Barton, Alison Etheridge,
Amandine Véber, Jerome Kelleher



Hitch-hiking in space

Genetic diversity only increases weakly with population size



Maynard Smith & Haigh (1974): in large populations, hitch-hiking is the main cause of coalescence

Probability of coalescence $\sim \exp(-2rt) \sim \exp(-2 \frac{t}{s} \log(4Ns)) \sim (4Ns)^{-2r/s}$

With spatial structure:

- spread is slower, allowing more time for recombination
- multiple founder events during spread of the new allele cause extra coalescence
- coalescence may be at the original mutation *or* at some local founder event

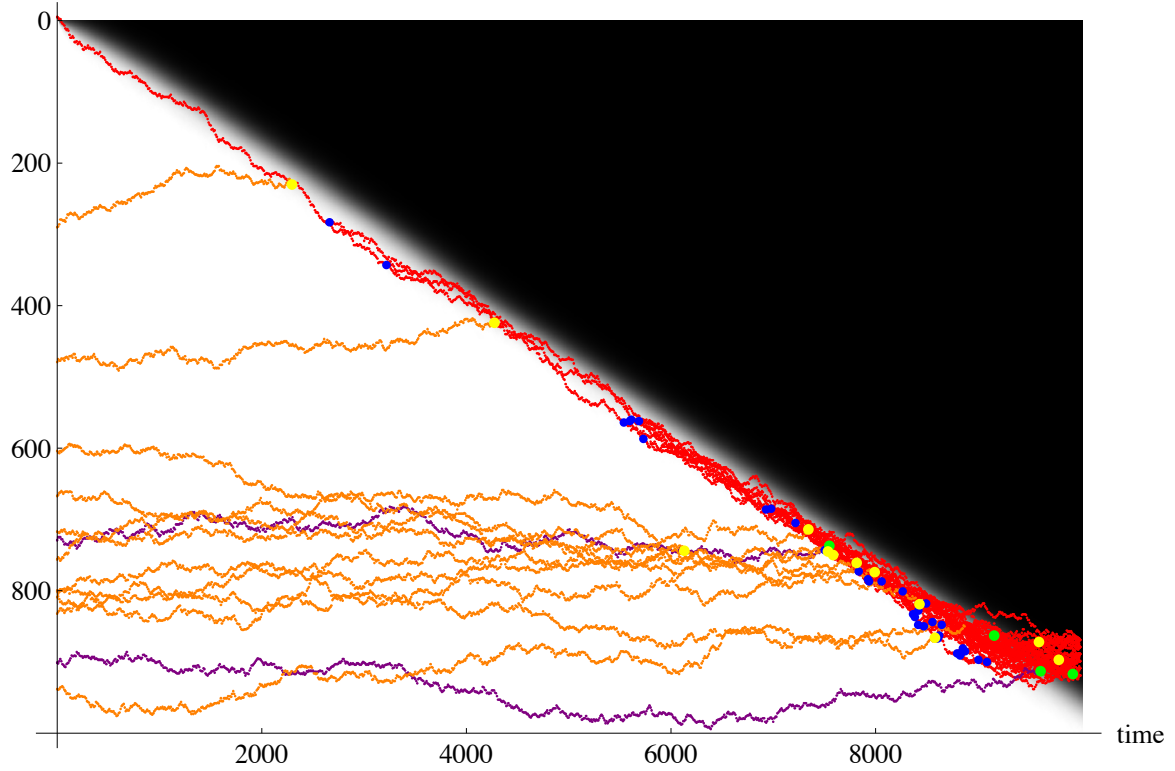
Hitch-hiking in one dimension

$s = 0.01$, $m = 0.5$, $N = 10^5$; $c \sim \sqrt{2ms} = 0.1$ demes per generation

red: selected locus; purple: 0.01cM; orange: 0.02cM

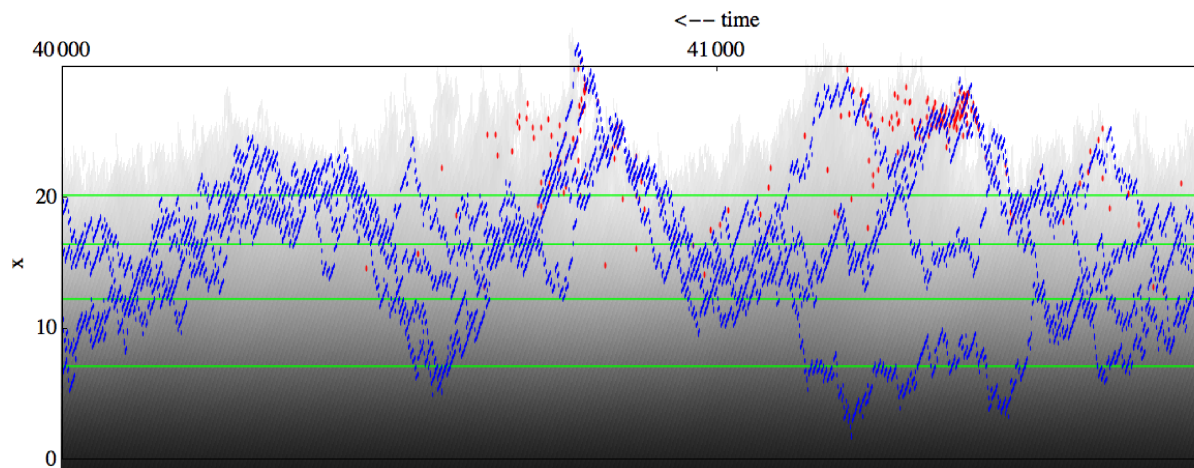
blue: coalescence; green: recombination within 0 - 0.01cM; yellow: recombination within 0.01 - 0.02cM

distance

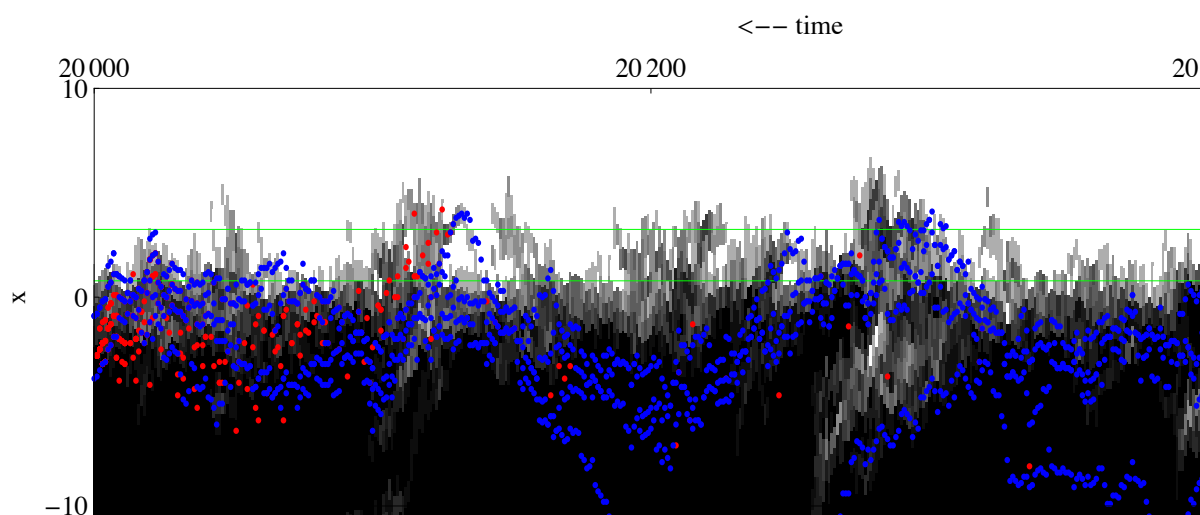


Fluctuations in one dimension

$N = 10^6$, $s = 0.05$, $m = 0.25$. **Green** : $p = 0.1, 0.01, 0.001, 0.0001$; **blue** : ancestral lineages; **red** : coalescence



$N = 100$, $s = 0.05$, $m = 0.25$. **Green**: $p=0.5, 0.2$; **blue**: ancestral lineages; **red**: coalescence



Cline shape in 1D

The Fisher-KPP equation has a *family* of travelling-wave solutions, $p[x - ct]$, $c \geq \sigma \sqrt{2s}$, *pulled* by reproduction at the tip

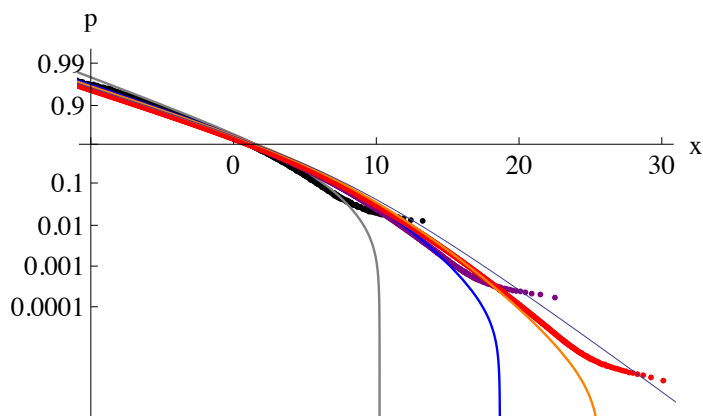
$$\frac{\partial p}{\partial t} = \frac{\sigma^2}{2} \frac{\partial^2 p}{\partial x^2} + sp(1-p) \quad (1)$$

Fluctuations at the tip slow down the wave to just *below* the minimum: $c \sim c_\infty \left(1 - \frac{A}{\log(B\rho\sigma\sqrt{s})^2} \right)$ (Brunet &

Derrida, 1997)

$A \sim 6.8$, $B \sim 24.8$; ρ is population density

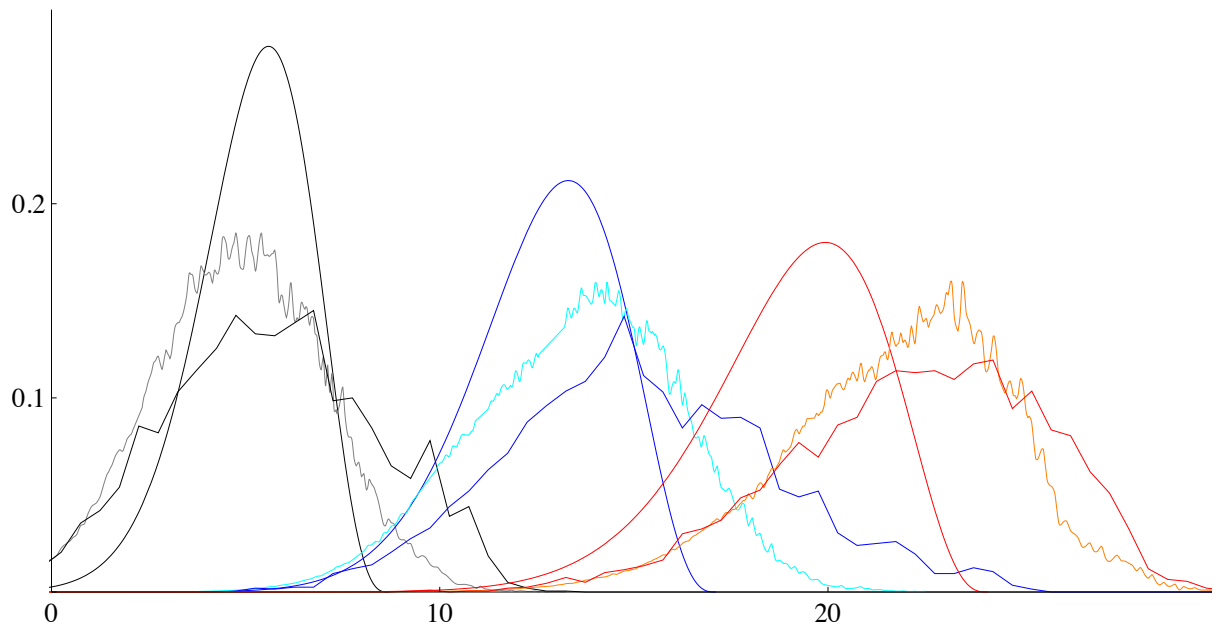
The cline is close to the deterministic shape, *given* reduced speed: $N = 100, 10^4, 10^6$



Predicting coalescence in 1D

Ignoring fluctuations, ancestors are distributed as $f = Cp^2 e^{2cx/\sigma^2}$ (Hallatschek & Nelson, 2008)

Coalescence occurs at rate $\int \frac{f^2}{Np} dx$

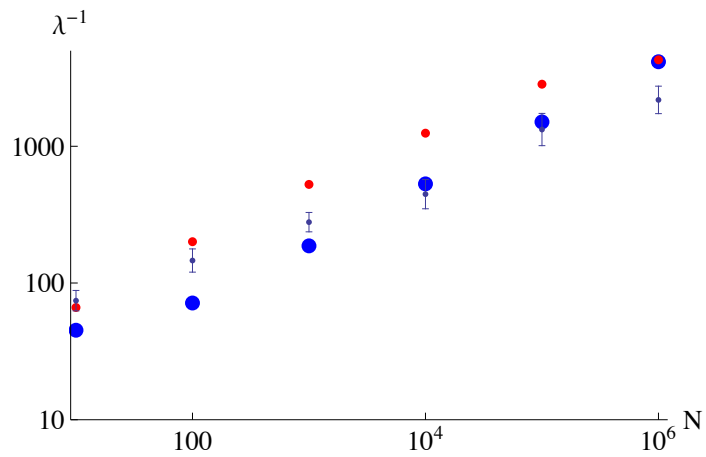


Rate of coalescence in 1D

The inverse rate of coalescence within the front is plotted against N .

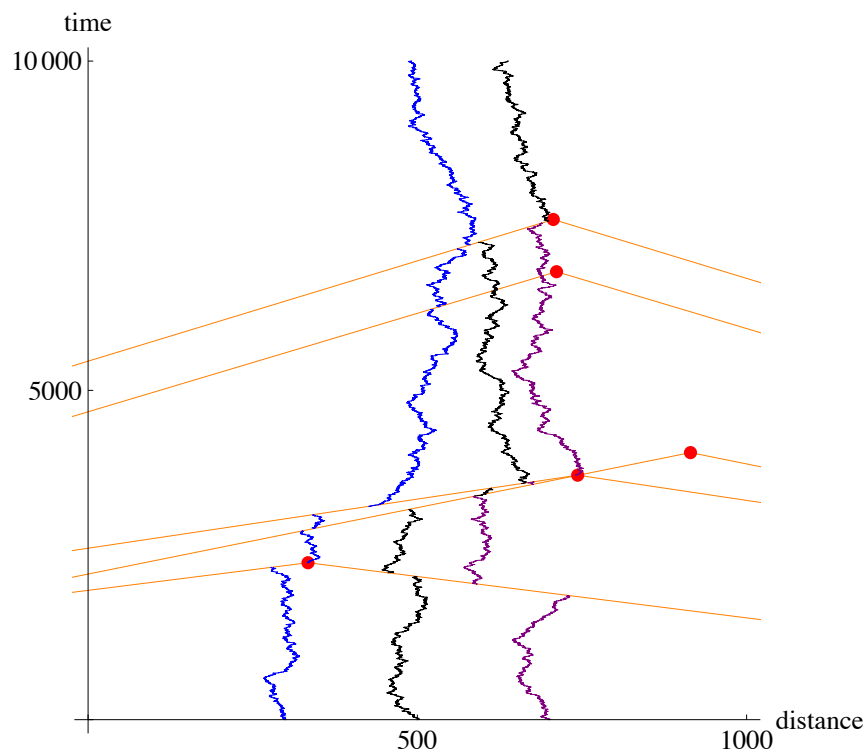
Large blue dots: *a priori* predictions, from the Fisher-KPP equation with $c \sim c_\infty \left(1 - \frac{A}{\log(B\rho\sigma\sqrt{s})^2} \right)$.

Small red dots: $\frac{f^2}{N\rho}$, based on the observed distribution of ancestors, f .



Coalescence in 1D

The effect if random sweeps; $\sigma^2 = 1$; the map locations, relative to the focal locus, are $-0.90cM$, $+0.80cM$, $-0.77cM$, $+0.83cM$, $-0.19cM$.



Random movement due to recurrent sweeps

Sweeps cause a random *jump* of $ct \sim \frac{c}{r}$. So, $\sigma_{\text{eff}}^2 = 2 \Lambda \left(\frac{c}{r}\right)^2$. Since $c \sim \sigma \sqrt{2s}$, $\frac{\sigma_{\text{eff}}^2}{\sigma^2} = \frac{4\Lambda s}{r^2}$

With tight linkage, and a finite range, L , the lineage may trace back to the original mutation

Averaging over a long genetic map ($\frac{RL}{c} \gg 1$) $\frac{\sigma_{\text{eff}}^2}{\sigma^2} = \frac{4}{3} \frac{L}{\ell} \frac{\Lambda}{R}$ where $\ell = \frac{\sigma}{\sqrt{2s}}$.

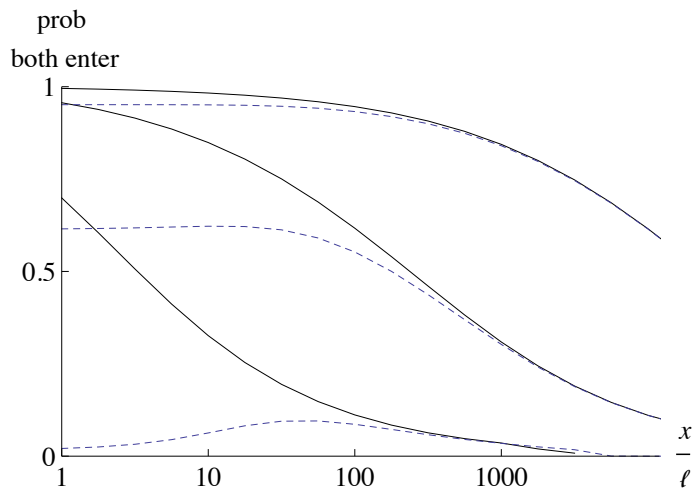
Even if $\frac{\Lambda}{R} \sim 0.001$, hitch-hiking can still be more important than diffusion if $L \gg 10^3 \ell$, say.

Probability of coalescence

The chance that two lineages x_1, x_2 from the front will hit at t_1, t_2 and coalesce within it is:

$$\frac{\lambda}{\lambda + 2r} \int_0^\infty \int_0^\infty \phi[t_1, x_1] \phi[t_2, x_2] e^{-r|t_2-t_1|} dt_1 dt_2 \quad \phi = \frac{x}{\sigma \sqrt{2\pi t^3}} \text{Exp}\left[-\frac{(x-ct)^2}{2\sigma^2 t}\right] \quad (2)$$

$\frac{r}{s} = 0.01, 0.1, 1$ (top to bottom); solid curves $\Delta x = 0$, dashed $\Delta x = 10\ell$



Long-term N_e

$\frac{1}{N_e}$ is defined as the rate of coalescence of randomly placed lineages

$$\frac{1}{N_e} = \frac{\Lambda}{L} \int_{-\infty}^{\infty} \frac{\lambda}{\lambda + 2r} \exp\left(-\frac{r|x|}{c}\right) dx = \Lambda \frac{\lambda}{\lambda + 2r} \frac{2c}{rL} \quad (3)$$

With tight linkage, the genes may trace back to the original mutation. Averaging over the genetic map, length R :

$$\frac{1}{N_e} = 2 \frac{\Lambda}{R} \frac{c}{L} g\left[\frac{L\lambda}{c}\right] \quad (4)$$

where $g[\theta]$ tends to $\log(4e) \sim 2.39$ for small θ , and to $2(\log(\frac{\theta}{2}) + \gamma - 1)$ for large θ , where $\gamma=0.577\dots$

This is insensitive to $L\lambda/c$, and is $\sim \Lambda/(RT)$ where $T \sim L/c$

Comparing with classical hitch-hiking: assume the genes are y apart when they hit the front, and that $\lambda \gg r$

$$\frac{\lambda}{\lambda + 2r} e^{-ry/c} \sim e^{-ry/c} > e^{-2(r/s)\log(2Ns)}$$

Map distance $\frac{c}{y}$ is affected, rather than $\frac{s}{\log(2Ns)}$; this is *smaller* because sweeps take longer to fix in a one-dimensional population

** Most movement may be due to hitch-hiking (i.e., $\sigma_{\text{eff}}^2 > \sigma^2$) even if sweeps cause little coalescence

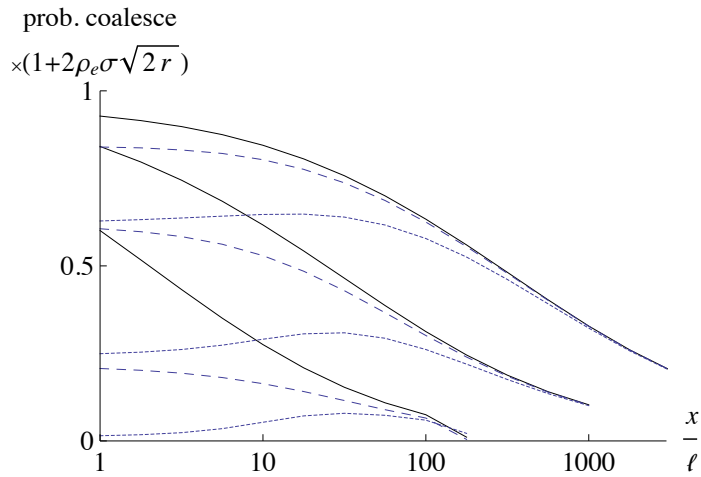
** Sweeps may still leave a strong local signature

Sweeps in two dimensions

Sweeps cause random jumps, as in one dimension: $\frac{\sigma_{\text{eff}}^2}{\sigma^2} = \frac{2}{3} \frac{L}{\ell} \frac{\Delta}{R}$

If genes trapped within the wavefront diffuse *along* it, then coalescence occurs as if in one dimension.

$\frac{r}{s} = 0.001, 0.01, 0.1$ (top to bottom); solid curves $\Delta x, \Delta y = 0$, long dashed curves $\Delta x=10\ell, \Delta y = 0$; dotted curves $\Delta x = 0, \Delta y = 10\ell$.



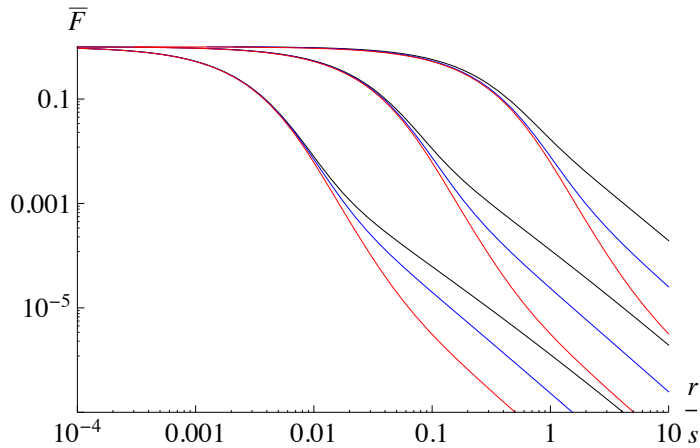
Sweeps in two dimensions

Averaging over randomly located genes:

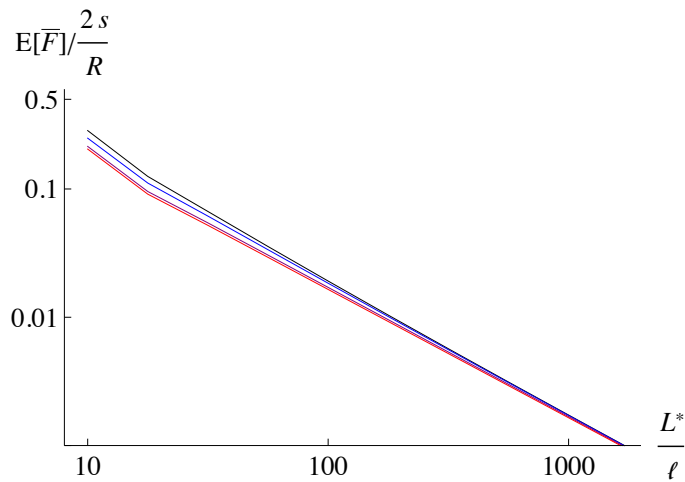
$$\Delta \left(\frac{2}{L} \right)^2 \int_0^\infty \int_0^\infty e^{-r\Delta x/c} \frac{\exp\left[-\frac{\Delta y}{\sigma} \sqrt{2r}\right]}{1 + 2\rho_e \sigma \sqrt{2r}} d\Delta x d\Delta y = \frac{4\Delta}{L^2} \frac{\sigma c}{r\sqrt{2r}} \frac{1}{1 + 2\rho_e \sigma \sqrt{2r}} \quad (5)$$

The probability that a sweep causes two genes randomly placed in a disk of radius L^* around the origin is insensitive to $\rho_e \sigma \sqrt{2s}$

$L^*/\ell = \mathcal{L} = 1000, 100, 10$ (left to right). Within each set, $2\rho_e \sigma \sqrt{2s} = 1, 10, 100$ (black, blue, red).

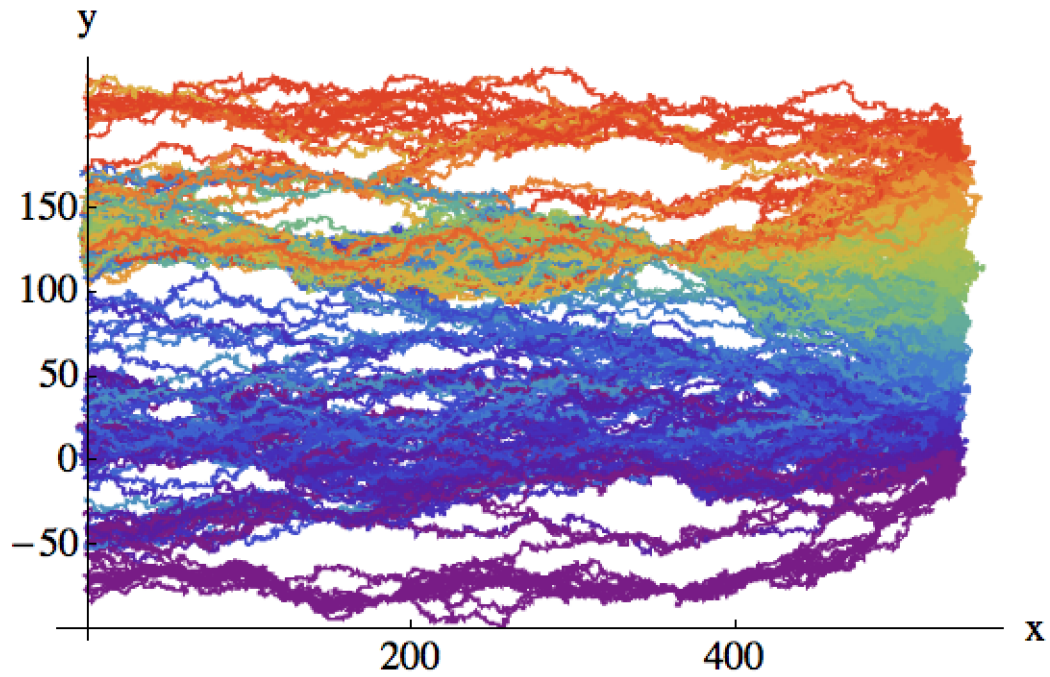


Averaging over the genetic map, $E[\bar{F}] \sim \frac{1.85c}{L^*R} = \frac{1.85}{RT}$: genes are simply drawn back to coalesce at the origin



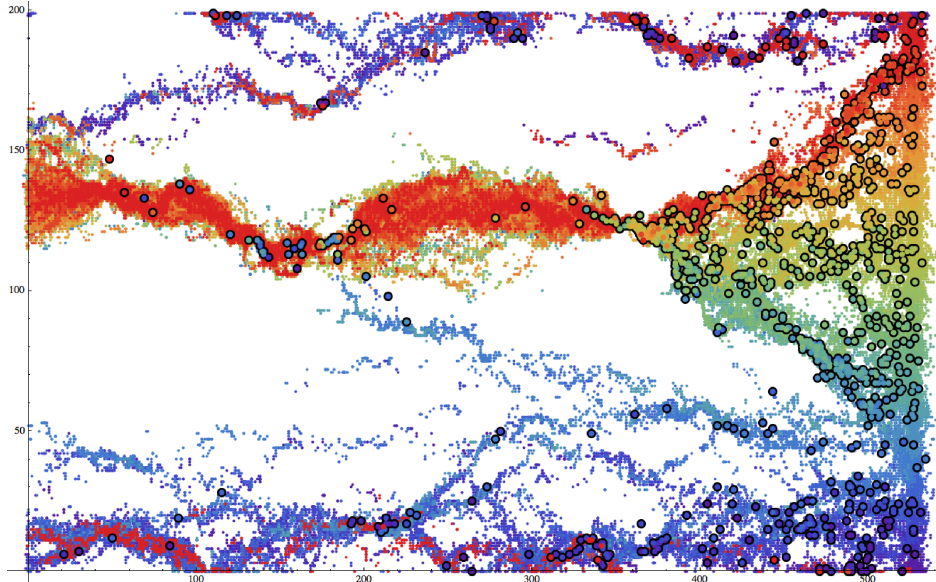
Fluctuations in two dimensions

Ancestral lineages diffuse transversely at $\sim \sigma^2$. $N = 100$, $s = 0.05$, $m = 0.25$, 4000 generations

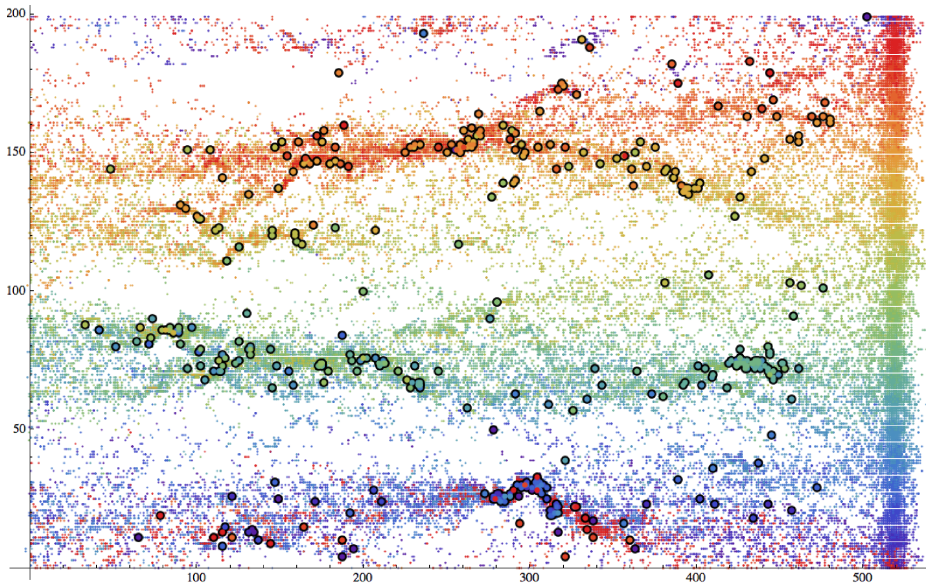


Fluctuations in two dimensions

However, coalescence events are strongly clustered:



Ancestral lineages diffuse transversely at $\sim \sigma^2$. $N = 10^6$, $s = 0.05$, $m = 0.25$, 4000 generations



Fluctuations in two dimensions

How can we describe the aggregation of ancestry?

Is the net rate of coalescence due to sweeps *faster* than in panmixis? NO

Can we detect the signature of *local* sweeps? YES