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 ~ \( \exp(-2rt) \sim \exp\left(-\frac{r}{4} \log(4Ns)\right) \sim (4Ns)^{-2r/s} \)

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

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

- **red**: selected locus
- **purple**: 0.01cM
- **orange**: 0.02cM
- **blue**: coalescence
- **green**: recombination within 0 - 0.01cM
- **yellow**: recombination within 0.01 - 0.02cM

\[ c ~ 2 ms = 0.1 \text{ demes per generation} \]

Distance vs. time graph showing multiple curves of different colors indicating the spread of hitch-hiking effects over time and 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, \( p(x - ct), c \geq \sigma \sqrt{2} s \), pulled by reproduction at the tip

\[
\frac{\partial p}{\partial t} = \frac{\sigma^2}{2} \frac{\partial^2 p}{\partial x^2} + sp(1-p) \tag{1}
\]

Fluctuations at the tip slow down the wave to just below the minimum: \( c - c_p \left( 1 - \frac{A}{\log(B/\sigma \sqrt{\epsilon})} \right) \) (Brunet & Derrida, 1997)

\( A \approx 6.8, B \approx 24.8; \rho \) 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 \( f = C p^2 e^{2\sqrt{x^2}} \) (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 $c \sim c_\infty \left(1 - \frac{A}{\log(B + \sqrt{A})}\right)$.

Small red dots: $\frac{\beta}{Np}$, 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 $ct - \frac{c}{t}$. So, $\sigma_{eff}^2 = 2 \Lambda \left(\frac{c}{t}\right)^2$. Since $c \sim \sigma \sqrt{2s}$, $\frac{\sigma_{eff}^2}{\sigma^2} = \frac{4\Lambda s}{\sigma^2}$.

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

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

Even if $\frac{L}{R} \sim 0.001$, 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:

\[
\phi = \frac{\lambda}{\lambda + 2x} \int_0^{t_2} \int_0^{t_1} \phi(t_1, x_1) \phi(t_2, x_2) e^{-\rho(t_2-t_1)} dt_1 dt_2
\]

\[
\phi = \frac{x}{\sigma \sqrt{2\pi t^3}} \exp \left[ -\frac{(x - \sigma t)^2}{2\sigma^2 t} \right]
\]

\[\tau_s = 0.01, 0.1, 1\] (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

$$\frac{1}{N_e} = \frac{\Lambda}{L} \int_{-\infty}^{0} \frac{\lambda}{\lambda + 2r} \exp\left(-\frac{r}{c} \left| x \right| \right) \, dx = \frac{\Lambda}{\lambda + 2r} \frac{2c}{xL}$$

(3)

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

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

(4)

where $g[\theta]$ tends to $\log(4\epsilon) \sim 2.39$ for small $\theta$, and to $2 \left( \log\left(\frac{\theta}{2}\right) + \gamma - 1 \right)$ for large $\theta$, where $\gamma \approx 0.577$...

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 $\lambda \gg r$

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

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

** Most movement may be due to hitch-hiking (i.e., $\sigma_{eff}^2 > \sigma^2$) 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: \( \frac{\partial \mu}{\partial t} = \frac{2}{3} \frac{r}{L} \Delta \)

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

\( \xi = 0.001, 0.01, 0.1 \) (top to bottom): solid curves \( \Delta x, \Delta y = 0 \), long dashed curves \( \Delta x = 10t, \Delta y = 0 \); dotted curves \( \Delta x = 0, \Delta y = 10t \).

\begin{align*}
\text{prob. coalesce} & \propto (1 + 2 \rho_c \sigma \sqrt{2r}) \\
\end{align*}

\( r_s = 0.001, 0.01, 0.1 \)
Sweeps in two dimensions

Averaging over randomly located genes:

\[
\Lambda \left( \frac{2}{L} \right)^2 \int_{0}^{\infty} \int_{0}^{\infty} e^{-r\Delta x/c} \exp \left[ -\frac{\Delta y}{\sigma} \sqrt{2 r} \right] \frac{d\Delta x \, d\Delta y}{1 + 2 \rho \sigma \sqrt{2 r}} = \frac{4 \Lambda}{L^2} \frac{\sigma c}{\sqrt{2 r}} \frac{1}{1 + 2 \rho \sigma \sqrt{2 r}} \quad (5)
\]

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

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

Averaging over the genetic map, \(E[F] = \frac{1.85 \times c}{L^* \ell} = \frac{1.85}{RR} : \) 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