

This is the equation describing the distribution of allele frequency across geographic space (Slatkin 73, eq. 11):

$$A(\xi) = -2 + 3 \tanh^2(\xi/2 + \tanh^{-1}(\sqrt{2/3})) \quad (1)$$

where

$$\xi = x/l_c \quad (2)$$

where

$$l_c = \sigma/\sqrt{s} \quad (3)$$

the characteristic length.

$$A = p - q = p - 1 + p = 2p - 1$$

Which means:

$$p(\xi) = \frac{A(\xi) + 1}{2} \quad (4)$$

If x is negative:

$$p = \frac{3}{2}(1 - \tanh^2(-\xi/2 + \tanh^{-1}(\sqrt{2/3}))) \quad (5)$$

0.0.1 Approximations of movement and recom. probs in tails

Furthermore, if x (or ξ) is really negative:

$$p = \frac{3}{2}(1 - \tanh^2(-\xi/2)) \quad (6)$$

Let $-\xi/2 = z$:

$$\tanh(z) = \frac{\exp(z) - \exp(-z)}{\exp(-z) + \exp(z)} \quad (7)$$

Divide through by $\exp(z)$:

$$= \frac{1 - \exp(-2z)}{\exp(-2z) + 1} \quad (8)$$

$$= \frac{1}{\exp(-2z) + 1} - \frac{\exp(-2z)}{\exp(-2z) + 1} \quad (9)$$

Since $-z$ is actually positive, the denominator is approximately $\exp(-2z)$. So:

$$= \frac{1}{\exp(-2z)} - \frac{\exp(-2z)}{\exp(-2z)} \quad (10)$$

$$= \frac{1}{\exp(-2z)} - 1 \quad (11)$$

$$= \exp(2z) - 1 \quad (12)$$

$$\tanh(z) \approx \exp(2z) - 1 \quad (13)$$

Which means that:

$$\tanh^2(z) \approx 1 - 2\exp(2z) + \exp(4z) \approx 1 - 2\exp(2z) \quad (14)$$

since $\exp(4z)$ goes to 0 faster than $\exp(2z)$

Sub into equation for p :

$$p \approx 3/2(1 - 1 + 2\exp(2z)) = 3\exp(2z)p \approx 6\exp(2z)/2 = 3\exp(2z) \quad (15)$$

$$(16)$$

Which makes the derivative:

$$\frac{dp}{dz} = 6\exp(2z) \quad (17)$$

Making

$$\frac{dp}{dz}/p = 2 \quad (18)$$

A constant drift to the right in the tail.

So that in unscaled units the ‘drift’ towards the right-side of the hybrid zone in the tails is

$$c = 2/l_c \quad (19)$$

That means that in the tails the change in position per generation is distributed as:

$$N(c, \sigma) \quad (20)$$

An approximate probability of the probability we’ve not recombined off Assuming that we can ignore the variance $c \gg \sigma$ (?) then if we are on the wrong background we move right at constant rare give(19) (at least in the tail). Assuming that if we recombine before we make it to the center of the hybrid zone, that we automatically recombine onto the good background. We sample a lineage a geographic distance L away from the hybrid zone, at a locus a genetic distance r

from selected locus. Then if τ is the time that the zone establishes, the probability that our lineage stays on the wrong background is

$$\approx \exp(-Lr/c), \text{ for } L/c < \tau \quad (21)$$

$$0, \text{ otherwise} \quad (22)$$

$$(23)$$

Taken from Barton and Etheridge spatial hitchhiking paper page 25-26:

The probability that the time for a Brownian motion with drift c hits zero for the first time at time t starting from x is

$$\phi(t, x) = \frac{x}{\sigma\sqrt{(2\pi t^3)}} \exp\left(-\frac{(x - ct)^2}{2\sigma^2 t}\right) \quad (24)$$

see e.g. Cox and Miller (1977) p.221, Equation 74 link . Notice that, at least if $x \gg \sigma$, this is essentially a Gaussian density, centred on $t = x/c$, the time since the wave was at the current location of the lineage, with standard deviation σ/c .

In combination with the approx. linear drift 19 we could use this to obtain an approximate distribution on the time it takes to reach the cline center. We could try to approximate prob of not recombining off bad background before we reach hybrid zone center. Assuming that if we recombine, that we'll fall off the bad background. The probability that we are still on the bad background by the time we reach hybrid zone center, over a time τ , is

$$\approx \int_0^\tau \phi(x, t) e^{-rt} dt \quad (25)$$

0.1 Covariance between ancestry in neutral non-equib. zone

The probability of recombining at time t is:

$$r e^{-rt}$$

After which point the two lineages are independent with variance (and covariance) of positions after time $\tau - t$ is:

$$(\tau - t)\sigma^2$$

We can take the integral over t to get:

$$\text{Var}(X) = \int (\tau - t) r e^{-rt} dt$$

Where X is the position of a lineage at time τ .

Now suppose that a lineage sampled at time 0 recombines at time t when the lineage is at position Z . At time τ the two lineages after recombination are at positions $x + Z$ and $y + Z$, where x and y are the displacement during time $(\tau - t)$.

$$\text{Cov}(x + Z, y + Z) = \text{Cov}(x, Z) + \text{Cov}(y, Z) + \text{Cov}(x, y) + \text{Cov}(Z, Z)$$

which comes down to:

$$\text{Cov}(Z, Z) = \text{Var}(Z, Z) = \sigma^2 t$$

We then integrate across t to get:

$$\int \sigma^2 t r e^{-rt} dt$$

Which gives us:

$$\sigma^2 \left(\tau - \frac{1 - e^{-r\tau}}{r} \right)$$

We now consider the case in which recombination does not occur in time τ . This happens with probability $e^{-r\tau}$. Since under this scenario, $X = Y$, $\text{Cov}(X, Y) = \text{Var}(X) = \sigma\tau$. This gives us a covariance of:

$$\sigma\tau e^{-r\tau} + \sigma^2 \left(\tau - \frac{1 - e^{-r\tau}}{r} \right)$$

Now, let us take a look at the covariance between X and Y , where H_X and H_Y are indicator variables such that $H_X = 1$ if X is on the RHS of the HZ, and $H_X = 0$ if X is on the LHS of the HZ. Suppose that we start with a lineage sampled on the LHS of the HZ, a distance L to the left of position 0 ($-L$). Under this scenario:

$$\begin{aligned} \text{Cov}(H_X, H_Y) &= \mathbb{E}[H_X H_Y] - \mathbb{E}[H_X] \mathbb{E}[H_Y] \\ &= \Pr(H_X = 1, H_Y = 1) - \mathbb{E}[H_X]^2 \end{aligned}$$

Since $\mathbb{E}[H_X] = \mathbb{E}[H_Y]$.

$$\mathbb{E}[H_X] = \Pr(H_X = 1) = \int_L^\infty x \sim \mathcal{N}(0, \sigma\sqrt{t}) dx$$

where x is the distance moved from starting point L . If recombination happens at time t at position $-L + z$ (i.e. lineage has moved distance z when recombination occurs)

$$\Pr(H_X = 1, H_Y = 1)$$

$$= \int_0^\tau r e^{-rt} \int_{-\infty}^\infty z \sim \mathcal{N}(0, \sigma\sqrt{t}) \int_{L-z}^\infty \int (x \sim \mathcal{N}(0, \sigma\sqrt{\tau-t}))(y \sim \mathcal{N}(0, \sigma\sqrt{\tau-t})) dx dy dz dt$$

This integral doesn't converge numerically in R or mathematica. We can simplify the expression to a single integral by expressing the normal distributions as a bivariate normal with vac-cov matrix:

$$\begin{pmatrix} \tau\sigma^2 & t\sigma^2 \\ t\sigma^2 & \tau\sigma^2 \end{pmatrix}$$

This reduces the quadruple integral to:

$$\int_0^\tau r\tau e^{-r\tau\frac{t}{\tau}} \int_L^\infty \int_L^\infty f_t(x, y) dx dy dt \quad (26)$$

Where

$$f(x, y) = \frac{1}{2\pi\tau\sigma^2\sqrt{1-(\frac{t}{\tau})^2}} \exp\left(-\frac{x^2 - 2\frac{t}{\tau}xy - y^2}{2\tau\sigma^2(1-(\frac{t}{\tau})^2)}\right)$$

Equation 26 is then

$$e^{-rt} \int_L^\infty \int_L^\infty f_t(x, y) dx dy|_{t=0}^{t=\tau} + \int_0^\tau e^{-r\tau t} \int_L^\infty \int_L^\infty \frac{\partial f_t}{\partial t} dy dx dt \quad (27)$$

We can calculate the double integral in the first term of equation 27 by interpreting that $t = 0$ means that recombination has happened at time 0 and the probabilities of both lineages moving L units are independent. Similarly, when $t = 1$, recombination does not happen until time τ and so we only need the probability that one lineage has moved L units. This gives us.

$$\left(1 - \Phi\left(\frac{L}{\sqrt{\tau}\sigma}\right)\right)^2 - e^{-r\tau} \left(1 - \Phi\left(\frac{L}{\sqrt{\tau}\sigma}\right)\right) \quad (28)$$

To solve the second term of equation 27, we use the fact that

$$\frac{\partial f_t}{\partial t} = \frac{\partial^2 f_t}{\partial x \partial y}$$

Which gives us

$$\begin{aligned} & \int_0^\tau e^{-rt} f_t(L, L) dt \\ &= \int_0^\tau e^{-rt} \frac{1}{2\pi\tau\sigma^2\sqrt{1-(\frac{t}{\tau})^2}} \exp\left(-\frac{L^2(\frac{t}{\tau}-1)}{\tau\sigma^2(1-(\frac{t}{\tau})^2)}\right) dt \end{aligned}$$

This integral converges with no trouble in R. Finally, since:

$$\text{Cov}(X, Y) = (1 - e^{-r\tau})\mathbb{E}[XY|\text{no rec.}] - (1 - e^{-r\tau})\mathbb{E}[X|\text{no rec.}]^2 + e^{-r\tau} \text{Cov}(X, X|\text{rec.})$$

We assemble all components of this to get:

$$\begin{aligned} \text{Cov}(X, Y) = & \int_0^\tau e^{-rt} \frac{1}{2\pi\tau\sigma^2\sqrt{1-(\frac{t}{\tau})^2}} \exp\left(-\frac{L^2(\frac{t}{\tau}-1)}{\tau\sigma^2(1-(\frac{t}{\tau})^2)}\right) dt \\ & + \left(1 - \Phi\left(\frac{L}{\sqrt{\tau}\sigma}\right)\right)^2 - e^{-r\tau} \left(1 - \Phi\left(\frac{L}{\sqrt{\tau}\sigma}\right)\right) \\ & - (1 - e^{-r\tau}) \left(1 - \Phi\left(\frac{L}{\sigma\sqrt{\tau}}\right)\right) \\ & + e^{-r\tau} \left(1 - \Phi\left(\frac{L}{\sigma\sqrt{\tau}}\right)\right) \Phi\left(\frac{L}{\sigma\sqrt{\tau}}\right) \end{aligned}$$

0.2 With selection:

Now consider two linked loci on a 'bad' background. If recomb. occurs on the LHS, one of the loci will end up on the 'correct' background, and the other will stay on the 'bad' one (remember factor of 2!). If, on the other hand, recomb happens on the RHS, both loci will be on what is now the 'correct' background.

On a stable cline, the distance traveled in time t becomes

$$L_t \sim \mathcal{N}(tc, t\sigma^2)$$

If recombination occurs on the LHS and locus 1(y_1) recombines onto the 'good' bg, then the distribution of locus y_1 is $\mathcal{N}(-tc, t\sigma^2)$ and the distribution of y_2 is $\mathcal{N}(tc, t\sigma^2)$

Not quite sure about how to get to a covariance with conditioning here....

Similar logic can be applied to the binary scenario. Is this on the right track??