1 Background selection approximations

The strength of background selection acting on neutral site ν is (where the sum x is over all linked conserved sites):

$$B(\nu) = \exp\left[-\sum_x \frac{u}{t(1+r_{x,\nu}\rho)^2}\right]$$

We use $\rho = (1-t)/t$ from Nordbord *et al.* 1996. Hudson and Kaplan 1995 used $\rho = 1/t$.

1.1 Contribution of contiguous conserved block to sum

For a contiguous conserved block consisting of bases $z = 1..n$, with a fixed recombination rate per base of q, and recombination distance d (between ν and the nearest base of the block):

$$r_{z,\nu} = d + (z-1)r$$

The contribution of this contigous block to the exponentiated sum can be approximated by integrating with respect to z over the block:

$$\begin{aligned}
& \sum_{z=1}^{n} \frac{u}{t(1+(d+zq)\rho)^2} \\
\approx\ & \frac{u}{t}\int_1^n \frac{dz}{(1+(d+(z-1)q)\rho)^2} \\
=\ & \frac{u}{t}\left[\frac{-t}{q(1-t)}\frac{1}{1+(d+(z-1)q)\rho}\right]_1^n \\
=\ & \frac{-u}{q(1-t)}\left[\frac{1}{1+(d+q(n-1))\rho} - \frac{1}{1+d\rho}\right] \\
=\ & \frac{u}{q(1-t)}\left[\frac{1}{1+r_{near,\nu}\rho} - \frac{1}{1+r_{far,\nu}\rho}\right]
\end{aligned}$$

Where r_{near} and r_{far} are the recombination distances of the near and far ends of the conserved block from the neutral site ν.

1.2 Contribution of remaining conserved sites to sum

Let m be the remaining recombination distance on the chromosome. Let c be the number of remaining conserved bases on the chromosome. Let r_ν be

the distance between the position ν that B is being evaluated at and our current position in the sum.

If we assume that conserved sites are evenly distributed along the remainder of the recombination distance, a conserved base occurs every m/c Morgans, and the sum over the remaining conserved sites $x = 1..c$ can be approximated as:

$$\begin{aligned} & \frac{u}{t}\sum_{x=1}^{c}\frac{1}{(1+\rho(r_\nu+\frac{m}{c}x))^2} \\ \approx & \frac{u}{t}\int_1^c \frac{dx}{(1+\rho(r_\nu+\frac{m}{c}x))^2} \\ = & \frac{u}{t}\frac{-ct}{m(1-t)}\left[\frac{1}{1+\rho(r_\nu+\frac{m}{c}x)}\right]_{x=1}^{x=c} \\ = & \frac{uc}{m(1-t)}\left[\frac{1}{1+\rho(r_\nu+\frac{m}{c})}-\frac{1}{1+\rho(r_\nu+m)}\right] \end{aligned}$$

We can also consider upper and lower bounds on the approximation. The greatest possible contribution would come if all of the sites were at the position r_ν. This gives us an upper bound on the remaining portion of the sum:

$$\frac{uc}{t(1+r_\nu\rho)^2}$$

The lowest possible contribution to the sum would occur if all sites are at the position $r_n u + m$. This gives us a lower bound on the remaining portion of the sum:

$$\frac{uc}{t(1+(r_\nu+m)\rho)^2}$$

A conservative approach towards approximating the sum is to terminate the summation when the remaining upper bound drops below a threshold, and to add in the remaining contribution using the uniformly distributed conserved site approximation described above.

1.3 Selection coefficient distribution

When using a distribution of selection coefficients, rather than a point estimate, the summation is instead:

$$\sum_x \int_0^1 \frac{uf(t)dt}{t(1+\rho r_{x,\nu})^2}$$

We can still approximate the sum over contiguous blocks with the integral approximation described above:

$$\frac{u}{r}\int_0^1 \frac{f(t)}{(1-t)}\Big[\frac{1}{1+\rho r_{near,\nu}} - \frac{1}{1+\rho r_{far,\nu}}\Big]dt$$

We evaluate integrals over t numerically and store results in a two dimensional table with indexes corresponding to r_{near} and $r_{len} = r_{far} - r_{near}$ values. When computing B-values and derivatives, fast estimates of the integrals are generated using bilinear interpolation from the nearest table entries.

1.4 Derivative of B

We want to calculate the derivative of B with respect to ν so that we can perform quadratic interpolation. We define $R(x)$ to be the recombination map position of conserved site x.

$$\begin{aligned}\frac{dB}{d\nu} &= \int_0^1 \frac{d}{d\nu}\Big[-\sum_x \frac{f(t)}{t(1+|R(x)-\nu|\rho)^2}\Big]Bdt \\ &= 2uB\int_0^1 \Big[\frac{\rho f(t)}{t}\sum_{x\in R(x)\le\nu}\frac{1}{(1+(\nu-R(x))\rho)^3} - \sum_{x\in R(x)>\nu}\frac{1}{(1+(R(x)-\nu)\rho)^3}\Big]dt\end{aligned}$$

We assume that blocks are split into two parts when they span the site ν and define $\delta_{x,\nu}$ as:

$$\delta_{x,\nu} = \begin{cases} -1 & \text{if } R(x) > \nu \\ 1 & \text{otherwise} \end{cases}$$

The derivative is now:

$$\frac{dB}{d\nu} = 2uB\int_0^1 \Big[\frac{\rho f(t)}{t}\sum_x \frac{\delta_{x,\nu}}{(1+r_{x,\nu}\rho)^3}\Big]dt$$

For sums of conserved bases within contiguous blocks and fixed recombination rates we can use an integral approximation similar to the one used above.

$$\begin{aligned}
& \sum_{z=1}^{n} \frac{1}{(1+(d+(z-1)q)\rho)^3} \\
\approx & \int_1^n \frac{dz}{(1+(d+(z-1)q)\rho)^3} \\
= & \frac{-1}{2q\rho}\Big[\frac{1}{(1+(d+(z-1)q)\rho)^2}\Big]_{z=1}^{z=n} \\
= & \frac{1}{2q\rho}\Big[\frac{1}{(1+r_{near,\nu}\rho)^2} - \frac{1}{(1+r_{far,\nu}\rho)^2}\Big]
\end{aligned}$$

Plugging this back into the derivative above (where the sum is now over conserved **blocks** rather than sites, and $r_{x_n,\nu}$ and $r_{x_f,\nu}$ are the recombination distances between ν and the near and far ends of conserved block x respectively:

$$\frac{dB}{d\nu} \approx uB\sum_x \frac{\delta_{x,\nu}}{q_x}\int_0^1 \Big[\frac{f(t)}{t(1+r_{x_n,\nu}\rho)^2} - \frac{f(t)}{t(1+r_{x_f,\nu}\rho)^2}\Big]dt$$

1.5 Second Derivative of B

$$\begin{aligned}
\frac{d^2B}{d\nu^2} &= \Big(\frac{dB}{d\nu}\Big)^2/B + 2uB\int_0^1 \frac{\rho f(t)}{t}\Big[\sum_{x\in R(x)\le\nu}\frac{-3\rho}{(1+(\nu-R(x))\rho)^4} + \sum_{x\in R(x)>\nu}\frac{3\rho}{(1+(R(x)-\nu)\rho)^4}\Big]dt \\
&= \Big(\frac{dB}{d\nu}\Big)^2/B - 6uB\int_0^1 \frac{\rho^2 f(t)}{t}\Big[\sum_x \frac{\delta_{x,\nu}}{(1+r_{x,\nu}\rho)^4}\Big]dt
\end{aligned}$$

For sums of conserved bases within contiguous blocks and fixed recombination rates we can use an integral approximation similar to the one already used for both the function and its first derivative.

$$\sum_{z=1}^{n} \frac{1}{(1+(d+(z-1)q)\rho)^4}$$

$$
\begin{aligned}
&\approx \int_1^n \frac{dz}{(1+(d+(z-1)q)\rho)^4} \\
&= \frac{-1}{3q\rho}\left[\frac{1}{(1+(d+(z-1)q)\rho)^3}\right]_{z=1}^{z=n} \\
&= \frac{1}{3q\rho}\left[\frac{1}{(1+r_{near,\nu}\rho)^3} - \frac{1}{(1+r_{far,\nu}\rho)^3}\right]
\end{aligned}
$$

Plugging this back into the second derivative above (where the sum is now over conserved **blocks** rather than sites, and $r_{x_n,\nu}$ and $r_{x_f,\nu}$ are the recombination distances between ν and the near and far ends of conserved block x respectively:

$$
\frac{d^2B}{d\nu^2} \approx \left(\frac{dB}{d\nu}\right)^2/B - 2uB\sum_x \frac{\delta_{x,\nu}}{q_x}\int_0^1 \left[\frac{\rho f(t)}{t(1+r_{x_n,\nu}\rho)^3} - \frac{\rho f(t)}{t(1+r_{x_f,\nu}\rho)^3}\right]dt
$$

1.6 Quadratic extrapolation

Using the first and second derivatives of $B(\nu)$ we can fit a quadratic to each positon we evaluate B at in order to estimate where the next position that we should evaluate B lies.

Consider a position ν_0 for which we know $B(\nu_0)$ and the first and second derivatives. We want to find a quadtratic $f(\nu)$ that passes through the same point, and has the same first and second derivatives at ν_0. The coefficients for this quadratic can be found as follows:

$$
f(\nu) = a\nu^2 + b\nu + c
$$

$$
\begin{aligned}
B''(\nu_0) &= f''(\nu_0) \\
B''(\nu_0) &= 2a \\
a &= \frac{1}{2}B''(\nu_0)
\end{aligned}
$$

$$
\begin{aligned}
B'(\nu_0) &= f'(\nu_0) \\
B'(\nu_0) &= 2a\nu_0 + b \\
B'(\nu_0) &= B''(\nu_0)\nu_0 + b \\
b &= B'(\nu_0) - B''(\nu_0)\nu_0
\end{aligned}
$$

$$
\begin{aligned}
B(\nu_0) &= f(\nu_0) \\
B(\nu_0) &= a\nu_0^2 + b\nu_0 + c \\
B(\nu_0) &= \frac{1}{2}B''(\nu_0)\nu_0^2 - B''(\nu_0)\nu_0^2 + B'(\nu_0)\nu_0 + c \\
c &= \frac{1}{2}B''(\nu_0)\nu_0^2 - B'(\nu_0)\nu_0 + B(\nu_0)
\end{aligned}
$$

Now we can find the next position to evaluate B at by extrapolate out from ν using our quadratic. Specifically, we want to find positive value κ such that $|f(\nu_0 + \kappa) - f(\nu_0)| = \epsilon$, where ϵ is our desired precision.

We can set the quadratic equal to $B(\nu_0) + / - \epsilon$ and solve for ν using the quadratic equation.

1.7 Truncated exponential distribution

One of the selection coefficient distributions we use is a truncated exponential distribution:

$$f(t) = Ce^{-\lambda t} \text{ for } a \leq t \leq b$$

$$f(t) = 0 \text{ otherwise}$$

We need to choose λ such that the mean of the truncated distribution is our desired selection coefficient mean, and C such that the integral over the distribution is 1.

$$
\begin{aligned}
1 &= \int_a^b Ce^{-\lambda t}dt \\
1 &= \left[\frac{-C}{\lambda}e^{-\lambda t}\right]_a^b \\
1 &= \frac{C}{\lambda}e^{-\lambda a} - \frac{C}{\lambda}e^{-\lambda b} \\
C &= \frac{\lambda}{e^{-\lambda a} - e^{-\lambda b}}
\end{aligned}
$$

We can use an s transform (see Drake *et al.* 1967, pp. 98-101) in order to calculate the expected value of the truncated distribution.

$$f_s^T(s) = E(e^{-st})$$

$$
\begin{aligned}
&= \int_{-\infty}^{\infty} Ce^{-st}e^{-\lambda t}dt \\
&= \int_{a}^{b} Ce^{-(s+\lambda)t}dt \\
&= \frac{-C}{s+\lambda}e^{-(s+\lambda)t}\Big|_{a}^{b} \\
&= \frac{C}{s+\lambda}[e^{-(s+\lambda)a} - e^{-(s+\lambda)b}]
\end{aligned}
$$

$$
\begin{aligned}
E[t] &= \left[\frac{df_t^T(s)}{dt}\right]_{s=0} \\
&= \frac{C(e^{-\lambda a} - e^{-\lambda b})}{\lambda^2} + \frac{ae^{-\lambda a} - be^{-\lambda b}}{\lambda} \\
&= \frac{1 + ae^{-\lambda a} - be^{-\lambda b}}{\lambda}
\end{aligned}
$$

This will still need to be solved numerically in order to obtain the appropriate lambda for the desired mean value.

2 Likelihood model

Under the Jukes-Cantor model, the probability of observing a mutational change along a branch i is:

$$p_i = \frac{3}{4}\left(1 - \exp\left[-\frac{4}{3}\beta_i\mu\right]\right)$$

Where β_i is the length of branch i in generations, and μ is the per-generation, per-nucleotide mutation rate.

If we use the Kimura two parameter model, there are separate probabilities for observing transitions (I) and transversions (V) as follows:

$$
\begin{aligned}
p_{I(i)} &= \frac{1}{4} - \frac{1}{2}\exp\left[-2(\mu_I + \mu_V)\beta_i\right] + \frac{1}{4}\exp\left[-4\mu_V\beta_i\right] \\
p_{V(i)} &= \frac{1}{2} - \frac{1}{2}\exp\left[-4\mu_V\beta_i\right]
\end{aligned}
$$

Under the kimura model the probability of an observed column, assuming no double substitutions is then:

$$\pi = \prod_i \left(p_{I(i)}\right)^{\delta_{I(i)}} \left(\frac{1}{2}p_{V(i)}\right)^{\delta_{V(i)}} \left(1 - p_{I(i)} - p_{V(i)}\right)^{(1-\delta_{I(i)}-\delta_{V(i)})}$$

where the product is over branches i and $\delta_{I(i)} = 1$ if the column represents a transition substitution and is assigned to branch i and $\delta_{I(i)} = 0$ otherwise. Likewise $\delta_{V(i)} = 1$ if the colulmn represents a transversion substitution assigned to branch i, and is 0 otherwise. The same formula applies for conserved alignment columns, in which case $\delta_{I(i)} = \delta_{V(i)} = 0$ for every branch.

Similarly, under the Jukes-Cantor model the probability of an observed column is:

$$\pi = \prod_i \left(\frac{1}{3}p_i\right)^{\delta_i} (1 - p_i)^{(1-\delta_i)}$$

In some cases we want to incorperate the probability that the observed alignment column is due to substitution on two branches. This is particularly important for the HG and CG columns, which, for certain parameter values, are more likely to be due to double substitutions than single substitutions on the HG or CG branches. Under the Jukes-Cantor and Kimura models, where forward and back substitutions occur at equal rates, tha same formulas as above can be used, with additional terms for each mutually exclusive single or double substitution class. For example, an HG column can be due to a single substitution on the HG branch or double substitution on the following pairs of branches: H+G, HCG+C. The probability of observing this column under the Jukes Cantor model is then:

$$\begin{aligned}
\pi_{HG} = & \frac{1}{3}(p_{HG}) \prod_{i \notin (HG)} (1 - p_i) \\
& + \frac{1}{9}(p_H)(p_G) \prod_{i \notin \{H,G\}} (1 - p_i) \\
& + \frac{1}{9}(p_{HCG})(p_C) \prod_{i \notin \{HCG,C\}} (1 - p_i)
\end{aligned}$$

$$LL = \sum_{B,\tau} n_{B,\tau} \log(\pi_\tau)$$

where $n_{B,\tau}$ is the number of filtered columns of type τ in bin B, and π_τ is the probability of alignment column type as described above.

The deleterious mutation rates were fixed in order to calculate the initial set of B-values. The deleterious rate can still be incorperated into the likelihood model by defining B as a function of two deleterious rate rescaling parameters (r_{ex} and r_{nex}) and the exonic and non-exonic B-values:

$$B(r_{\text{ex}}, r_{\text{nex}}) = \exp\left[\log(B_{\text{ex}})r_{\text{ex}} + \log(B_{\text{nex}})r_{\text{nex}}\right]$$

So we can perform a gradient search of the likelihood surface we calculate the partial derivatives of the log-likelihood function w.r.t. each of the model parameters ($\bullet$):

$$\frac{\partial LL}{\partial \bullet} = \sum_{B,\tau} \frac{n_{B,\tau}}{\pi_\tau} \frac{\partial \pi_\tau}{\partial \bullet}$$

The partial derivatives of the column probabilities can be found using (by the product rule):

$$\begin{aligned}
\frac{\partial \pi_\tau}{\partial \bullet} &= \left(\sum_i \frac{\frac{\partial p_i^{\delta_i}(1-p_i)^{(1-\delta_i)}}{\partial \bullet}}{p_i^{\delta_i}(1-p_i)^{(1-\delta_i)}} \right) \pi_\tau \\
&= \left(\sum_{i=\tau} \frac{\frac{\partial p_i}{\partial \bullet}}{p} - \sum_{i \neq \tau} \frac{\frac{\partial p_i}{\partial \bullet}}{1-p_i} \right) \pi_\tau
\end{aligned}$$

And the partial derivatives of the observed branch-mutation probabilities are:

$$\frac{\partial p_i}{\partial \bullet} = \left(\frac{\partial \beta_i}{\partial \bullet}\mu + \frac{\partial \mu}{\partial \bullet}\beta_i \right) \exp\left(-\frac{4}{3}\beta_i \mu \right)$$

For all parameters, with the exception of μ, $\frac{\partial \mu}{\partial \bullet}$ is 0. The branch partial derivatives are zero w.r.t. μ, but must be non-zero w.r.t. at least a subset the deleterious mutation rate, effective population size, and speciation time parameters. These partial derivatives vary depending on the model.

2.1 HC model

We define the length of human and chimp branches as:

$$\begin{aligned}\beta_H &= \beta_C = 2N_{HC}B + T_{HC} \\ \beta_{H+C} &= 4N_{HC}B + 2T_{HC}\end{aligned}$$

The non-zero partial derivatives of the combined human+chimp branch are:

$$\begin{aligned}\frac{\partial \beta_{H+C}}{\partial N_{HC}} &= 4B \\ \frac{\partial \beta_{H+C}}{\partial T_{HC}} &= 2 \\ \frac{\partial \beta_{H+C}}{\partial r_-} &= 4N_{HC} \log(B_-) B\end{aligned}$$

2.2 HCM model

In the human, chimp, macaque model the definitions of the human and chimp branches and partial derivatives are the same as in the HC model.

We can define the length of the outgroup macaque branch as the sum of the right branch (leading from root to macaque) and left branches (leading from root to start of human and chimp branches). The length of the right side of the branch is $1.4(T_{HCM} + T_{HC}) + 2N_{HCM}B$, where the 1.4 is a correction for the accelerated mutation rate in old world monkeys. The length of the left side of the branch is $2N_{HCM}B + T_{HCM} - 2N_{HC}B$, so the combined length is:

$$\beta_M = 4N_{HCM}B + T_{HCM} + 1.4(T_{HCM} + T_{HC}) - 2N_{HC}B$$

The non-zero partial derivatives of this branch are:

$$\begin{aligned}\frac{\partial \beta_M}{\partial N_{HC}} &= -2B \\ \frac{\partial \beta_M}{\partial N_{HCM}} &= 4B\end{aligned}$$

$$\begin{aligned}
\frac{\partial \beta_M}{\partial T_{HC}} &= 1.4 \\
\frac{\partial \beta_M}{\partial T_{HCM}} &= 2.4 \\
\frac{\partial \beta_M}{\partial r_-} &= (4N_{HCM} - 2N_{HC}) \log(B_-) B
\end{aligned}$$

2.3 HCGOM model

Expected branch lengths for each site type are given by:

$$\begin{aligned}
\beta_H &= \beta_C = T_{HC} + 2BN_{HC} + \kappa_{HCG}\left(\frac{4}{3}BN_{HCG} - 2BN_{HC}\right) \\
\beta_{H+C} &= 2\beta_H \\
\beta_G &= T_{HC} + T_{HCG} + 2\left(1 - \frac{\kappa_{HCG}}{3}\right) BN_{HCG} \\
\beta_{HG} &= \beta_{CG} = \left(\frac{2}{3}\right) \kappa_{HCG} BN_{HCG} \\
\beta_{HC} &= T_{HCG} + (2BN_{HCG} - 2BN_{HC})(1 - \kappa_{HCG}) + \left(\frac{2}{3}\right) \kappa_{HCG} BN_{HCG} \\
\beta_{HCG} &= T_{HCGO} + 2BN_{HCGO} - 2BN_{HCG}\left(1 + \frac{k_{HCG}}{3}\right) \\
\beta_O &= T_{HC} + + T_{HCG} + T_{HCGO} + 2BN_{HCGO} \\
\beta_M &= 1.4\,(T_{HC} + T_{HCG} + T_{HCGO} + T_{HCGOM}) + T_{HCGOM} + 4BN_{HCGOM} - 2BN_{HCGO}
\end{aligned}$$

where $\kappa_{HCG} = \exp\left(-\frac{T_{HCG}}{2BN_{HC}}\right)$ is the probability that the human-chimpanzee coalescent predates the gorilla speciation.

$$\frac{\partial \kappa_{HCG}}{\partial N_{HC}} = \frac{T_{HCG}}{2B(N_{HC})^2} \kappa_{HCG}$$

$$
\begin{aligned}
\frac{\partial \kappa_{HCG}}{\partial T_{HCG}} &= -\frac{1}{2B(N_{HC})}\kappa_{HCG} \\
\frac{\partial \kappa_{HCG}}{\partial r_-} &= \frac{T_{HCG}\log(B_-)}{2N_{HC}B}\kappa_{HCG}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial \beta_H}{\partial N_{HC}} &= 2B + \frac{\partial \kappa_{HCG}}{\partial N_{HC}}\left(\frac{4}{3}BN_{HCG} - 2BN_{HC}\right) - \kappa_{HCG}2B \\
&= 2B\left(1 - \kappa_{HCG} + \frac{\partial \kappa_{HCG}}{\partial N_{HC}}\left[\frac{2}{3}N_{HCG} - N_{HC}\right]\right) \\
\frac{\partial \beta_H}{\partial N_{HCG}} &= \frac{4}{3}B\kappa_{HCG} \\
\frac{\partial \beta_H}{\partial T_{HC}} &= 1 \\
\frac{\partial \beta_H}{\partial T_{HCG}} &= \frac{\partial \kappa_{HCG}}{\partial T_{HCG}}\left(\frac{4}{3}BN_{HCG} - 2BN_{HC}\right) \\
\frac{\partial \beta_H}{\partial r_-} &= 2N_{HC}\log(B_-)B + \left(\frac{\partial \kappa_{HCG}}{\partial r_-} + \log(B_-)\kappa_{HCG}\right)\left(\frac{4}{3}BN_{HCG} - 2BN_{HC}\right)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial \beta_G}{\partial N_{HC}} &= -\frac{2}{3}BN_{HCG}\frac{\partial \kappa_{HCG}}{\partial N_{HC}} \\
\frac{\partial \beta_G}{\partial N_{HCG}} &= 2\left(1 - \frac{\kappa_{HCG}}{3}\right)B \\
\frac{\partial \beta_G}{\partial T_{HC}} &= 1 \\
\frac{\partial \beta_G}{\partial T_{HCG}} &= 1 - \frac{2}{3}BN_{HCG}\frac{\partial \kappa_{HCG}}{\partial T_{HCG}} \\
\frac{\partial \beta_G}{\partial r_-} &= 2\log(B_-)B\left(1 - \frac{\kappa_{HCG}}{3}\right)N_{HCG} - \frac{2}{3}BN_{HCG}\frac{\partial \kappa_{HCG}}{\partial r_-}
\end{aligned}
$$

$$
\frac{\partial \beta_{HC}}{\partial N_{HC}} = 2B\left(\kappa_{HCG} - 1 + \frac{\partial \kappa_{HCG}}{\partial N_{HC}}N_{HC} - \frac{2}{3}\frac{\partial \kappa_{HCG}}{\partial N_{HC}}N_{HCG}\right)
$$

$$
\begin{aligned}
\frac{\partial \beta_{HC}}{\partial N_{HCG}} &= 2B(1-\frac{2}{3}\kappa_{HCG}) \\
\frac{\partial \beta_{HC}}{\partial T_{HCG}} &= 1-2B\frac{\partial \kappa_{HCG}}{\partial T_{HCG}}\left(\frac{2}{3}N_{HCG}-N_{HC}\right) \\
\frac{\partial \beta_{HC}}{\partial r_{-}} &= 2B\log(B_{-})\left(N_{HCG}-N_{HC}-\frac{2}{3}N_{HCG}\kappa_{HCG}+N_{HC}\kappa_{HCG}\right)+ \\
&\quad 2B\frac{\partial \kappa_{HCG}}{\partial r_{-}}\left(N_{HC}-\frac{2}{3}N_{HCG}\right)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial \beta_{HG}}{\partial N_{HC}} &= \frac{2}{3}\frac{\partial \kappa_{HCG}}{\partial N_{HC}}BN_{HCG} \\
\frac{\partial \beta_{HG}}{\partial N_{HCG}} &= \frac{2}{3}\kappa_{HCG}B \\
\frac{\partial \beta_{HG}}{\partial T_{HCG}} &= \frac{2}{3}\frac{\partial \kappa_{HCG}}{\partial T_{HCG}}BN_{HCG} \\
\frac{\partial \beta_{HG}}{\partial r_{-}} &= \frac{2}{3}\left(\log(B_{-})\kappa_{HCG}+\frac{\partial \kappa_{HCG}}{\partial r_{-}}\right)BN_{HCG}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial \beta_{HCG}}{\partial N_{HC}} &= -\frac{2}{3}BN_{HCG}\frac{\partial \kappa_{HCG}}{\partial N_{HC}} \\
\frac{\partial \beta_{HCG}}{\partial N_{HCG}} &= -2B\left(1+\frac{\kappa_{HCG}}{3}\right) \\
\frac{\partial \beta_{HCG}}{\partial N_{HCGO}} &= 2B \\
\frac{\partial \beta_{HCG}}{\partial T_{HCG}} &= -\frac{2}{3}BN_{HCG}\frac{\partial \kappa_{HCG}}{\partial T_{HCG}} \\
\frac{\partial \beta_{HCG}}{\partial T_{HCGO}} &= 1 \\
\frac{\partial \beta_{HCG}}{\partial r_{-}} &= 2B\left(\log(B_{-})N_{HCGO}-\log(B_{-})N_{HCG}-\frac{\partial \kappa_{HCG}}{\partial r_{-}}N_{HCG}\right)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial \beta_O}{\partial N_{HCGO}} &= 2B \\
\frac{\partial \beta_O}{\partial T_{HC}} &= 1 \\
\frac{\partial \beta_O}{\partial T_{HCG}} &= 1 \\
\frac{\partial \beta_O}{\partial T_{HCGO}} &= 1 \\
\frac{\partial \beta_O}{\partial r_-} &= 2\log(B_-)BN_{HCGO}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial \beta_M}{\partial N_{HCGO}} &= -2B \\
\frac{\partial \beta_M}{\partial N_{HCGOM}} &= 4B \\
\frac{\partial \beta_M}{\partial T_{HC}} &= 1.4 \\
\frac{\partial \beta_M}{\partial T_{HCG}} &= 1.4 \\
\frac{\partial \beta_M}{\partial T_{HCGO}} &= 1.4 \\
\frac{\partial \beta_M}{\partial T_{HCGOM}} &= 2.4 \\
\frac{\partial \beta_M}{\partial r_-} &= 4B\log(B_-)N_{HCGOM} - 2B\log(B_-)N_{HCGO}
\end{aligned}
$$

2.4 Branch lengths

The probability of a coalescence in each generation is $\frac{n(n-1)}{4N}$ where n is the number of sampled copies.

The time till the first coalescence event in n sampled copies is exponentially distributed with mean $\frac{4N}{n(n-1)}$. We can denote the time at which the number of remaining branches in the gene tree is reduced from n to $n-1$ as as $T(n)$.

The probability that the human-chimp coalescent predates gorilla speciation is, from the exponential cumulative distribution function:

$$\kappa_{HCG} = P(T(2) \geq T_{HCG}) = exp(-T_{HCG}/2N_{HC})$$

2.4.1 H and C branches

Assume that the human-chimp coalescent predates gorilla speciation. Then there are three possible arrangements of the branches. The H and C branches may coalesce first, the H and G branches may coalesce first, or the C and G branches may coalesce first. The expected length of the H branch is dictated by these three mutually exclusive possibilities:

$$\begin{aligned} & \kappa_{HCG}\left[T_{HCG} + \frac{1}{3}\left(\frac{2}{3}N_{HCG}B\right) + \frac{1}{3}\left(\frac{2}{3}N_{HCG}B\right) + \frac{1}{3}\left(\frac{2}{3}N_{HCG}B + 2N_{HCG}B\right)\right] \\ = \; & \kappa_{HCG}\left[T_{HCG} + \frac{4}{3}N_{HCG}B\right] \end{aligned}$$

Now consider the contribution to the branch length from when the HC coalescent is less than T_{HCG}. We must integrate over the branch length probability distribution up to T_{HCG}:

$$\begin{aligned} & \int_{x=0}^{x=T_{HCG}} \frac{x}{2N_{HC}B} \exp\left(-\frac{x}{2N_{HC}B}\right) dx \\ = \; & \frac{1}{2N_{HC}B}\left[\exp\left(-\frac{x}{2N_{HC}B}\right)(2N_{HC}B)^2\left(-\frac{x}{2N_{HC}B} - 1\right)\right]_{x=0}^{x=T_{HCG}} \\ = \; & \left[\exp\left(-\frac{x}{2N_{HC}B}\right)(-x - 2N_{HC}B)\right]_{x=0}^{x=T_{HCG}} \\ = \; & \exp\left(-\frac{T_{HCG}}{N_{HC}B}\right)(-T_{HCG} - 2N_{HC}B) + 2N_{HC}B \\ = \; & 2N_{HC}B - \kappa_{HCG}T_{HCG} - \kappa_{HCG}2N_{HC}B \end{aligned}$$

Combining the above and adding in the branch length up till human-chimp speciation gives:

$$\begin{aligned}\beta_H = \beta_C &= T_{HC} + \kappa_{HCG} T_{HCG} + \kappa_{HCG} \frac{4}{3} N_{HCG} B \\ &\quad + 2N_{HC}B - \kappa_{HCG} T_{HCG} - \kappa_{HCG} 2N_{HC} B \\ &= T_{HC} + 2N_{HC}B - \kappa_{HCG} \left(\frac{4}{3} N_{HCG} B - 2N_{HC} B \right)\end{aligned}$$

2.4.2 G branch

If the human-chimp coalescent is older than the gorilla speciation event, then there are again three possible orderings of coalescent events. If the HG or CG branches coalesce first, then the expected length of the gorilla branch in the ancestral population is $\frac{2}{3} N_{HCG} B$, otherwise the expected length is $\frac{8}{3} N_{HCG} B$. Combining this with the possibility that the HC coalescent did not predate the gorilla speciation event, and the lineage specific divergence time $(T_{HC} + T_{HCG})$ gives the following:

$$\begin{aligned}\beta_{HCG} &= T_{HC} + T_{HCG} + (1 - \kappa_{HCG}) 2N_{HCG} B + \\ &\quad \kappa_{HCG} \left[\frac{2}{3} \left(\frac{2}{3} N_{HCG} B \right) + \frac{1}{3} \left(\frac{8}{3} N_{HCG} B \right) \right] \\ &= T_{HC} + T_{HCG} + 2N_{HCG} B \left(1 - \frac{\kappa_{HCG}}{3} \right)\end{aligned}$$

2.4.3 HC branch

The length of the HC branch is the time separating the human-chimp/gorilla coalescent and the human/chimp coalescent.

If the H/C coalescent does not predate gorilla speciation the HC branch length is the time to coalescence of HC and G in the HCG population $(T_{HCG} + 2N_{HCG})$ minus the time to H/C coalescence in the HC population. We must integrate over the H/C coalescent time distribution in order to find this difference:

$$\begin{aligned}&(1 - \kappa_{HCG}) (T_{HCG} + 2N_{HCG}) - \int_{x=0}^{x=T_{HCG}} \frac{x}{2N_{HC}B} \exp\left(-\frac{x}{2N_{HC}B} \right) dx \\ = \;& T_{HCG} + 2N_{HCG}B - 2N_{HC}B - 2\kappa_{HCG} N_{HCG} B + 2\kappa_{HCG} N_{HC} B \\ = \;& T_{HCG} + (2N_{HCG}B - 2N_{HC}B)(1 - \kappa_{HCG})\end{aligned}$$

If the H/C coalescent time is older than gorilla speciation, then only one of the three possible gene trees contains an HC branch, and the length of this branch is $2N_{HCG}B$, so:

$$\kappa_{HCG}\frac{2}{3}N_{HCG}B$$

Combining both of the above gives:

$$\beta_{HC} = T_{HCG} + (2N_{HCG}B - 2N_{HC}B)(1-\kappa_{HCG}) + \frac{2}{3}\kappa_{HCG}N_{HCG}B$$

Note that in the original paper draft we subtract out the "old" HC sites that predate gorilla speciation, so we omit the $\kappa_{HCG}\frac{2}{3}N_{HCG}B$ term from the formula.

2.4.4 HG and CG branches

HG and CG branches only occur when the H/C coalescent predates gorilla speciation. Each of these branches only only occur for a single of the three possible gene-tree topologies, and the branch length is always $2N_{HCG}B$. So we have:

$$\beta_{HG} = \beta_{CG} = \frac{2}{3}\kappa_{HCG}N_{HCG}B$$

2.4.5 HCG branch

The length of the HCG branch is the time separating gorilla and orang speciation T_{hcgo}, plus the time to coalescent in the ancestral HCGO population ($2N_{HCGO}B$), minus the length of the coalescent in the ancestral HCG population ($\kappa_{HCG}\left(\frac{8}{3}N_{HCG}B\right) + (1-\kappa_{HCG})2N_{HCG}B$):

$$\beta_{HCG} = T_{HCGO} + 2N_{HCGO}B - 2N_{HCG}B\left(1+\frac{\kappa_{HCG}}{3}\right)$$

2.4.6 O branch

The length of the O branch is simply the sum of interspeciation speciation times plus the expected coalescent time in the O population:

$$\beta_O = T_{HC} + T_{HCG} + T_{HCGO} + 2N_{HCGO}B$$

2.4.7 M branch

The length of the M branch is the length of the branch leading down to macaque plus the length of the branch leading down to the HCGO common ancestor (minus the HCGO coalescent) plus twice the coalescent time in the HCGOM commen ancestor. The macaque side of the branch is multiplied by 1.4 to account for the faster rate of evolution in old world monkeys.

$$\begin{aligned}\beta_M &= T_{HCGO} + 4N_{HCGOM}B - 2N_{HCGO}B \\ &\quad +1.4\,(T_{HC} + T_{HCG} + T_{HCGO} + T_{HCGOM})\end{aligned}$$