1. Birth-death-recombination process

Forwards in time, suppose that individuals divide and die independently at rates λ and μ, respectively, starting with an initial population of one at time 0. The associated genealogical tree is obtained by representing each individual as an edge, which splits when the individual divides, and terminates when the individual dies. Suppose also that each lineage undergoes recombination at rate ρ: in the genealogy, a second lineage is selected uniformly at random to merge with the recombining lineage (which inherits part of its genetic material to the left of the recombination breakpoint from one of the parents, and the rest from the other parent). Hence, just after the recombination event, the population size decreases by one. The process is run for a random time T^*_{or}, at which point an extant population is of size N is observed.

The total population size thus follows a birth-death process with birth rate λ and total death rate $\mu+\rho$. A known result is that the backwards-in-time equivalent obtained by conditioning on observing an extant population of size N at the present, and reversing time with respect to a uniform prior on the time to origin T_{or}, is a birth-death process with the birth and death rates swapped. We thus define the backwards-in-time birth-death-recombination process, in which individuals are added at rate μ, recombine (split) at rate ρ, and die at rate λ, starting with N individuals at time 0 and reaching extinction at the random time T_{or} in the past.

A uniform (improper) prior on T_{or} is a common choice (e.g. Aldous and Popovic, 2005; Wiuf, 2018), and after conditioning on observing the extant population size N, a proper distribution on T_{or} is obtained, with $T_{or} = T^*_{or}$ in distribution. Following Section 1.1.1 (BD tree paper), the posterior density of T_{or} in this case is given by

$$f_{T_{or}}(t) = \begin{cases} \dfrac{N\mu e^{(\mu+\rho-\lambda)t}\left[\frac{\mu}{\mu+\rho-\lambda}(e^{(\mu+\rho-\lambda)t}-1)\right]^{N-1}}{\left[1+\frac{\mu}{\mu+\rho-\lambda}(e^{(\mu+\rho-\lambda)t}-1)\right]^{N+1}} & \text{if } \mu+\rho > \lambda \\ \dfrac{N\lambda e^{(\lambda-\mu-\rho)t}\left[\frac{\lambda}{\lambda-\mu-\rho}(e^{(\lambda-\mu-\rho)t}-1)\right]^{N-1}}{\left[1+\frac{\lambda}{\lambda-\mu-\rho}(e^{(\lambda-\mu-\rho)t}-1)\right]^{N+1}} & \text{if } \lambda > \mu+\rho. \end{cases} \tag{1.1}$$

Assume that $\lambda > \mu+\rho$ from now on, for simplicity.

A realisation of this birth-death-recombination process describing the history of the full population (the *complete* process) can thus be obtained by starting with a sample of N individuals at the present time, and adding birth, death and recombination events until eventual extinction. The corresponding genealogy is constructed by selecting the lineages that undergo each event uniformly at random.

1.1. The reversed reconstructed process ($\mathbf{RRP}_\rho$).

The complete process traces out the history of the entire population, while the object of interest is often the genealogy of a sample of size $n < N$ obtained at the present time. We will consider the case of Bernoulli sampling: each extant lineage is sampled independently with fixed probability ψ. The process tracing out the genealogy of the sample backwards in time is the reversed reconstructed process (RRP); for the case of no recombination its properties have been considered extensively (e.g. Gernhard, 2008; Stadler, 2009; Wiuf, 2018; Ignatieva et al., 2020).

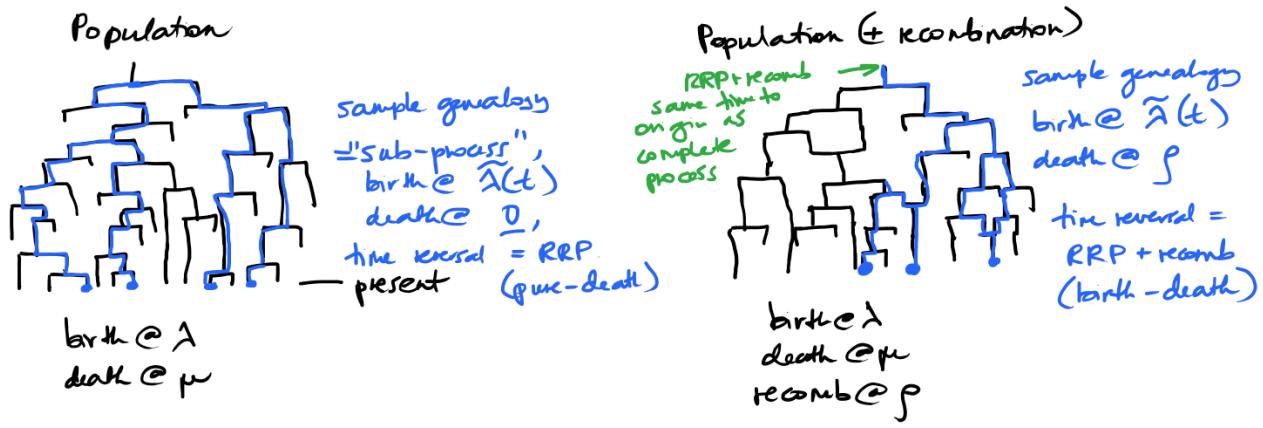
A realisation of this process can be obtained by taking a realisation of the complete process, and removing any non-sampled and extinct lineages; note that in the presence of recombination, the genealogy takes the shape of an ARG rather than a tree. However, this method of simulation is very inefficient, even for moderate sample sizes and reasonably small values of ψ, and does not allow for the properties of the RRP to be easily obtained. Thus, we next formulate the RRP explicitly as a birth-death process (which tracks only those events from the complete process that affect the sampled lineages).

Looking backwards from the present time 0, each *sampled* lineage splits into two independently at rate ρ. The death rate of each lineage is

$$m(t) = \lambda \cdot P_\psi(0,t) \tag{1.2}$$

$$= \frac{\psi\lambda e^{(\lambda-\mu)t}}{1+\frac{\psi\lambda}{\lambda-\mu}(e^{(\lambda-\mu)t}-1)}, \tag{1.3}$$

where $P_\psi(0,t)$ is the probability that the lineage has at least one descendant that does not undergo a death event, given by Yang and Rannala (1997). The RRP is defined as the birth-death process with $X(t)$ individuals at time $t \geq 0$ in the past, with birth rate ρ and inhomogeneous death rate $m(t)$, starting with $X(0) = n$ individuals and run until extinction. Note that when $\psi = 1$ and $\mu = 0$, we have $m(t) = \lambda$: in the absence of deaths and with full sampling, the RRP is equivalent to the complete process.



1.2. **Properties.** Define

$$\gamma(t) = \int_0^t (m(\tau) - \rho)d\tau = \log\left(1 + \frac{\psi\lambda}{\lambda - \mu}(e^{(\lambda-\mu)t} - 1)\right) - \rho t, \tag{1.4}$$

so that

$$e^{\gamma(t)} = \left(1 + \frac{\psi\lambda}{\lambda - \mu}(e^{(\lambda-\mu)t} - 1\right) e^{-\rho t} = \frac{\psi\lambda}{\lambda - \mu}e^{(\lambda-\mu-\rho)t} - \frac{\psi\lambda}{\lambda - \mu}e^{-\rho t} + e^{-\rho t}. \tag{1.5}$$

Following Bailey (1964, Section 9.3), let

$$A(t) = \int_0^t \rho e^{\gamma(\tau)}d\tau \tag{1.6}$$

$$= 1 - e^{\rho t} + \frac{\psi\lambda}{\lambda - \mu}\left(e^{\rho t} - 1\right) + \frac{\psi\lambda\rho}{(\lambda - \mu)(\lambda - \mu - \rho)}\left(e^{(\lambda-\mu-\rho)t} - 1\right), \tag{1.7}$$

and using the identity

$$\frac{\psi\lambda}{\lambda - \mu} + \frac{\psi\lambda\rho}{(\lambda - \mu)(\lambda - \mu - \rho)} = \frac{\psi\lambda}{\lambda - \mu - \rho}, \tag{1.8}$$

we have

$$e^{\gamma(t)} + A(t) = 1 + \frac{\psi\lambda}{\lambda - \mu - \rho}\left(e^{(\lambda-\mu-\rho)t} - 1\right). \tag{1.9}$$

Let

$$\alpha(t) = 1 - \frac{1}{e^{\gamma(t)} + A(t)}, \quad \beta(t) = 1 - \frac{e^{\gamma(t)}}{e^{\gamma(t)} + A(t)}.$$

1.2.1. *Time to origin.* Using classical results for birth-death processes (Bailey, 1964, Section 9.3, p. 110), the distribution of the time to origin $T_{or} = \inf\{t \geq 0 : X(t) = 0\}$ is given by

$$F_{or}(t) = \mathbb{P}(T_{or} < t) = [\alpha(t)]^n = \left(\frac{\frac{\psi\lambda}{\lambda-\mu-\rho}\left(e^{(\lambda-\mu-\rho)t} - 1\right)}{1 + \frac{\psi\lambda}{\lambda-\mu-\rho}\left(e^{(\lambda-\mu-\rho)t} - 1\right)}\right)^n. \tag{1.10}$$

Note that this is the same as the distribution of time to origin for the complete process, with density for $\psi = 1$ given by (1.1), as the root lineage of the RRP genealogy does not die until the the time of extinction of the complete process. This tends to 1 as $t \to \infty$ for $\lambda > \mu + \rho$, but the expectation of T_{or} is not finite.

1.2.2. *Time to GMRCA.* First passage time $T_{GMRCA} = \inf\{t \geq 0 : X(t) = 1 | X(0) = n\}$.

1.2.3. *Number of lineages through time.* The probability that at time t there are k lineages is given by (Bailey, 1964)

$$f_k(t) = \sum_{j=0}^{\min(k,n)} \binom{n}{j} \binom{n+k-j-1}{n-1} \alpha(t)^{n-j} \beta(t)^{k-j} (1-\alpha(t)-\beta(t))^j, \tag{1.11}$$

and the mean number of lineages at time t is

$$\mathbb{E}(X(t)) = ne^{-\gamma(t)} = \frac{ne^{\rho t}}{1 + \frac{\psi\lambda}{\lambda-\mu}\left(e^{(\lambda-\mu)t}-1\right)}$$

but want 1 to be an absorbing state

1.2.4. *Width of the ARG.* What is the expected maximum number of lineages?

1.2.5. *Number of recombination events.* What is the distribution of the number of recombination events?

2. Simulation

The RRP can be simulated in a straightforward manner through drawing exponentially distributed waiting times between events, and using time rescaling to obtain the event times of the RRP. Define the intensity

$$\Lambda_\psi(t) = \int_0^t (\rho + m(\tau))d\tau = \log\left(1 + \frac{\psi\lambda}{\lambda-\mu}(e^{(\lambda-\mu)t}-1)\right) + \rho t. \tag{2.1}$$

The inverse Λ^{-1} does not have a nice closed form for general values of the parameters, but can be computed numerically. The function Λ_ψ^{-1} is a time rescaling between the RRP with time-inhomogeneous total event rate, and a process where events happen at total rate n when there are n lineages, but the probabilities of each event *type* are time-dependent. Each event is a recombination with probability $\rho/\left(\rho + m(\Lambda_\psi^{-1}(y))\right)$ and a coalescence with probability $m(\Lambda_\psi^{-1}(y))/\left(\rho + m(\Lambda_\psi^{-1}(y))\right)$.

2.1. **Small ψ.** Considering the case when ψ is small,

$$\Lambda_\psi(t) = \frac{\psi\lambda}{\lambda-\mu}\left(e^{(\lambda-\mu)t}-1\right) + \rho t + \mathcal{O}(\psi^2),$$

using Taylor expansion about $\psi = 0$. Set

$$\frac{\psi\lambda}{\lambda-\mu}\left(e^{(\lambda-\mu)t}-1\right) + \rho t = y.$$

Then for $\rho > 0$,

$$t = \frac{\psi\lambda + (\lambda-\mu)y}{(\lambda-\mu)\rho} - \frac{\psi\lambda}{(\lambda-\mu)\rho}e^{(\lambda-\mu)t}.$$

This is of standard form: the solution to $x = a + be^{cx}$ is $x = a - \frac{1}{c}W(-bce^{ac})$, where $a, b \neq 0, c \neq 0$ are constants and W is the Lambert W function. Thus,

$$\Lambda_0^{-1}(y) \approx \widetilde{y} - W\left(\frac{\psi\lambda}{(\lambda-\mu)\rho}e^{\widetilde{y}}\right),$$

where $\widetilde{y} = \frac{\psi\lambda+(\lambda-\mu)y}{(\lambda-\mu)\rho}$. When t is very large, this approximation breaks down. Instead we have

$$\Lambda_\psi(t) \approx \log\left(\frac{\psi\lambda}{\lambda-\mu}e^{(\lambda-\mu)t}\right) + \rho t = \log\left(\frac{\psi\lambda}{\lambda-\mu}\right) + (\lambda-\mu+\rho)t, \tag{2.2}$$

giving

$$\Lambda_0^{-1}(y) \approx \frac{1}{\lambda-\mu+\rho}\left(y - \log\left(\frac{\psi\lambda}{\lambda-\mu}\right)\right)$$

There are thus two time regimes with a smooth transition between them. Can simulate the two regimes separately and combine in some way to get an approximation?

Algorithm 1 Simulating an ARG

Start with n lineages at time 0.
Set $Y = 0$.
While $n > 1$:

(1) Draw $\widetilde{Y} \sim \text{Exp}(n)$ and $Z \sim \text{Unif}(0, 1)$.
(2) Set $Y = Y + \widetilde{Y}$, rescale $t = \Lambda_\psi^{-1}(Y)$, and calculate $p = \rho/(\rho + m(t))$.
(3) At time t, if $Z < p$, select a lineage at random and split into two, set $n = n + 1$. Otherwise, select two lineages uniformly at random and merge together, set $n = n - 1$.

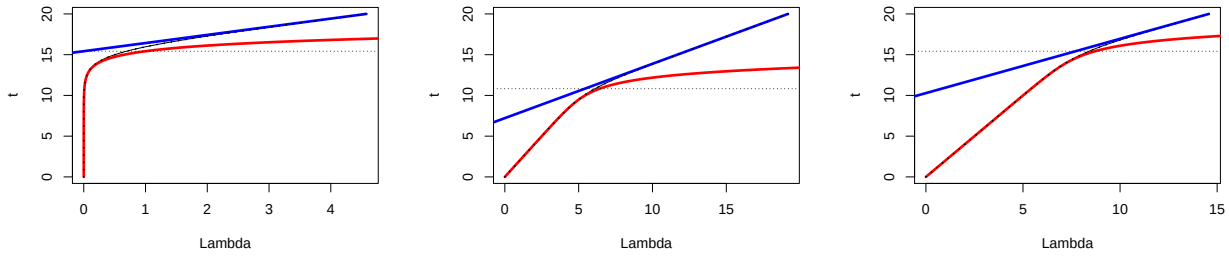


Figure 1. Left: $\rho = 0, \psi = 10^{-7}$, middle: $\rho = 0.5, \psi = 10^{-5}$, right: $\rho = 0.5, \psi = 10^{-7}$. Red and blue lines: small psi approximations (two regimes). Dotted line: actual relationship.

2.2. **Different time scaling.** Define the time transformation

$$g_1(x) = \log\left(1 + \frac{\psi\lambda}{\lambda - \mu}\left(e^{(\lambda-\mu)x} - 1\right)\right)$$

and

$$g_1^{-1}(x) = \frac{1}{\lambda - \mu}\log\left(1 + \frac{\lambda - \mu}{\psi\lambda}\left(e^x - 1\right)\right)$$

Then the death rate on this time scale is

$$m_1(x) = m(g_1^{-1}(x))\left|\frac{d}{dx}g_1^{-1}(x)\right| = 1$$

and the recombination rate is

$$\rho \cdot \left|\frac{d}{dx}g_1^{-1}(x)\right| = \frac{\rho\frac{1}{\lambda-\mu}\frac{\lambda-\mu}{\psi\lambda}e^x}{1 + \frac{\lambda-\mu}{\psi\lambda}\left(e^x - 1\right)} = \frac{\rho e^x}{\psi\lambda + (\lambda - \mu)\left(e^x - 1\right)}$$

$$= \frac{\rho}{(\lambda - \mu)\left[\frac{\psi\lambda}{\lambda-\mu}e^{-x} + 1 - e^{-x}\right]} = \frac{\rho}{\lambda - \mu} \cdot \frac{1}{1 - \left(1 - \frac{\psi\lambda}{\lambda-\mu}\right)e^{-x}}$$

The limit as $x \to \infty$ is $\frac{\rho}{\lambda-\mu}$ and as $x \to 0$ is $\frac{\rho}{\psi\lambda}$.

Consider the critical limit $\lambda - \mu \to 0$, so that the total population size is constant in expectation. The rescaled recombination rate becomes

$$\frac{\rho}{\psi\lambda}e^x.$$

Taking $\rho \to 0$ and $\psi \to 0$ so that $\frac{\rho}{\psi\lambda} \to \alpha$, the recombination rate is αe^x and the intensity is

$$\Lambda(t) = \int_0^t (1 + \alpha e^x)\, dx = t + \alpha e^t - \alpha$$

Then

$$y = \Lambda(t) = t + \alpha e^t - \alpha \implies t = y - \alpha e^t \implies \Lambda^{-1}(y) = \alpha + y - W(\alpha e^{\alpha+y})$$

where W is the Lambert W function.

3. Large population limit

Consider the expression for $\gamma(t)$ in (1.4). When t is small, Taylor expansion around 0 gives

$$\gamma(t) = (\lambda\psi - \rho)t + \mathcal{O}(t^2),$$

so the RRP behaves like a birth-death process with birth rate ρ and death rate $\psi\lambda$. When t is large, using $\log(1+x) = \log(x) + \log(1+1/x)$

$$\log\left(1 + \frac{\psi\lambda}{\lambda-\mu}\left(e^{(\lambda-\mu)t} - 1\right)\right) = \log\left(\frac{\psi\lambda}{\lambda-\mu}\right) + \log\left(e^{(\lambda-\mu)t} - 1\right) + \log\left(1 - \frac{1}{\frac{\psi\lambda}{\lambda-\mu}\left(e^{(\lambda-\mu)t} - 1\right)}\right) \tag{3.1}$$

$$\implies \gamma(t) \approx \log\left(\frac{\psi\lambda}{\lambda-\mu}\right) + (\lambda - \mu - \rho)t, \tag{3.2}$$

so the RRP behaves like a birth-death process with birth rate ρ and death rate $\lambda - \mu$.

References

Aldous, D. and Popovic, L. (2005). A critical branching process model for biodiversity. *Advances in Applied Probability*, **37**(4), 1094–1115.

Bailey, N. T. (1964). *The elements of stochastic processes with applications to the natural sciences*. Wiley.

Gernhard, T. (2008). The conditioned reconstructed process. *Journal of Theoretical Biology*, **253**(4), 769–778.

Ignatieva, A., Hein, J. and Jenkins, P. A. (2020). A characterisation of the reconstructed birth–death process through time rescaling. *Theoretical Population Biology*, **134**, 61–76.

Stadler, T. (2009). On incomplete sampling under birth–death models and connections to the sampling-based coalescent. *Journal of Theoretical Biology*, **261**(1), 58–66.

Wiuf, C. (2018). Some properties of the conditioned reconstructed process with Bernoulli sampling. *Theoretical population biology*, **122**, 36–45.

Yang, Z. and Rannala, B. (1997). Bayesian phylogenetic inference using DNA sequences: a Markov Chain Monte Carlo method. *Molecular Biology and Evolution*, **14**(7), 717–724.