(This is the standard coalescent simulation routine in BEAST, except that I modified it for an earliest possible time for the whole tree to coalesce by.)

Suppose $f : \mathbb{R} \to [0, \infty)$ is a function such that $f(t)$ is the effective population size of the population at time t, where the timescale is in backwards time. Define the intensity function $I : \mathbb{R} \to [0, \infty)$ as $I(t) = \int_0^t \frac{1}{f(s)} ds$, and suppose that I has a calculable inverse $I^{-1} : [0, \infty) \to \mathbb{R}$.

If there are N lineages at time t_0, then the probability density function for the time until the first coalescence is:

$$\begin{aligned} p(t) &= \frac{N(N-1)}{2f(t_0+t)} \exp\left(-\int_{t_0}^{t_0+t} \frac{N(N-1)}{2f(s)} ds \right) \\ &= \frac{N(N-1)}{2f(t_0+t)} \exp\left(\frac{N(N-1)}{2} (I(t_0) - I(t_0+t)) \right) \end{aligned}$$

and the cumulative density function of this is:

$$\begin{aligned} P(t) &= \int_0^t \frac{N(N-1)}{2f(t_0+r)} \exp\left(-\int_{t_0}^{t_0+r} \frac{N(N-1)}{2f(s)} ds \right) dr \\ &= 1 - \exp\left(\frac{N(N-1)}{2} (I(t_0) - I(t_0+t)) \right) \end{aligned}$$

Suppose that we know that all lineages must coalesce before time T (for example, because T is the known time of infection of an individual, and we are assuming transmission is a complete bottleneck). If we condition P on this, the CDF becomes:

$$Q(t) = \begin{cases} \frac{1-\exp\left(\frac{N(N-1)}{2}(I(t_0)-I(t_0+t))\right)}{1-\exp\left(\frac{N(N-1)}{2}(I(t_0)-I(T))\right)} & t < T - t_0 \\ 1 & t \geq T - t_0 \end{cases}$$

A sample from this distribution be taken by sampling a value U from the uniform distribution on [0,1], and then solving $Q(t) = U$ for t. If $D = 1 - \exp\left(\frac{N(N-1)}{2}(I(t_0) - I(T))\right)$:

$$
\begin{aligned}
U &= \frac{1-\exp\left(\frac{N(N-1)}{2}(I(t_0)-I(t_0+t))\right)}{D} \\
UD &= 1-\exp\left(\frac{N(N-1)}{2}(I(t_0)-I(t_0+t))\right) \\
\ln(1-UD) &= \frac{N(N-1)}{2}(I(t_0)-I(t_0+t)) \\
I(t_0)-\frac{2}{N(N-1)}\ln(1-UD) &= I(t_0+t) \\
I^{-1}\left(I(t_0)-\frac{2}{N(N-1)}\ln(1-UD)\right) &= t_0+t \\
I^{-1}\left(I(t_0)-\frac{2}{N(N-1)}\ln(1-UD)\right)-t_0 &= t
\end{aligned}
$$

So suppose we have a time T of first infection, and the times of N tips $t_1, \ldots, t_N$ with all $t_i < T$; there is one t_i for each sampling event for this host, and one for each other host it directly infects. We start with the smallest t_i and move along the timeline, simulating intervals. If there is only one lineage at time t then obviously there are no coalescences in the interval between t and the first t_i greater than t, or between t and T if there is no such t_i; in the first case we update the time to that t_i, and in the second we are finished. Otherwise, if there are $n \geq 2$ lineages at t then draw a coalescence time c using the procedure above. Let t_j be the smallest t_i with $t_j > t$. If $t + c < t_j$ then we coalesce two random lineages at $t + c$ and then update to simulating a new interval at $t + c$. If $t + c \geq t_j$ then we coalesce nothing and update to simulating a new interval at t_j, because the coalescence of the existing lineages would occur after a new tip was introduced and we need to recalculate at that point to account for this tip.

Once we've got to time T then we have a phylogeny that has coalesced N tips to a single lineage. It has a root branch length, which is the different between the time of root and T. We do this for every host, and then build the phylogeny from the transmission tree by connecting the root of each host's phylogeny to the tip of the phylogeny of its infector that corresponds to its infection, by a branch of this length.