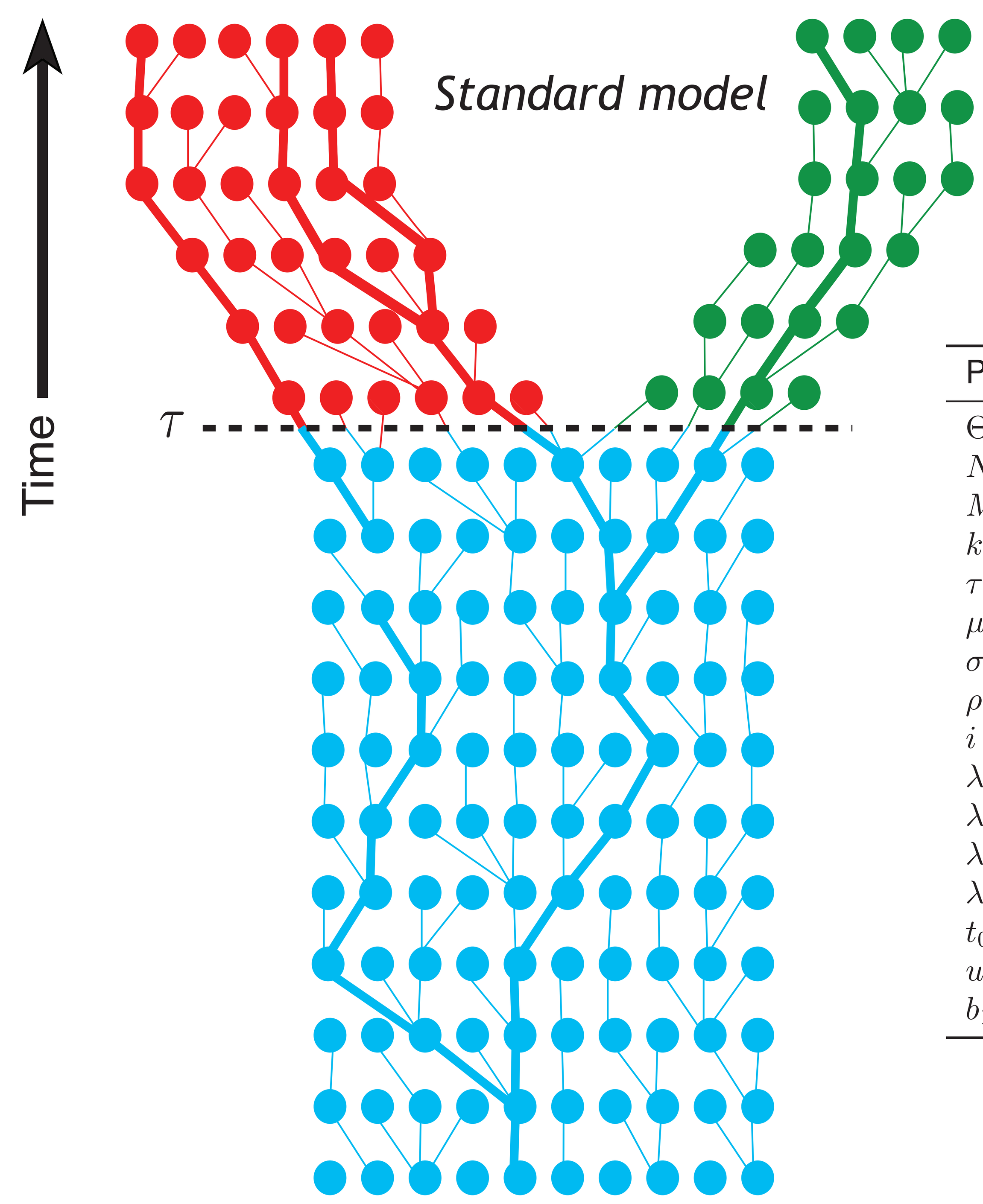
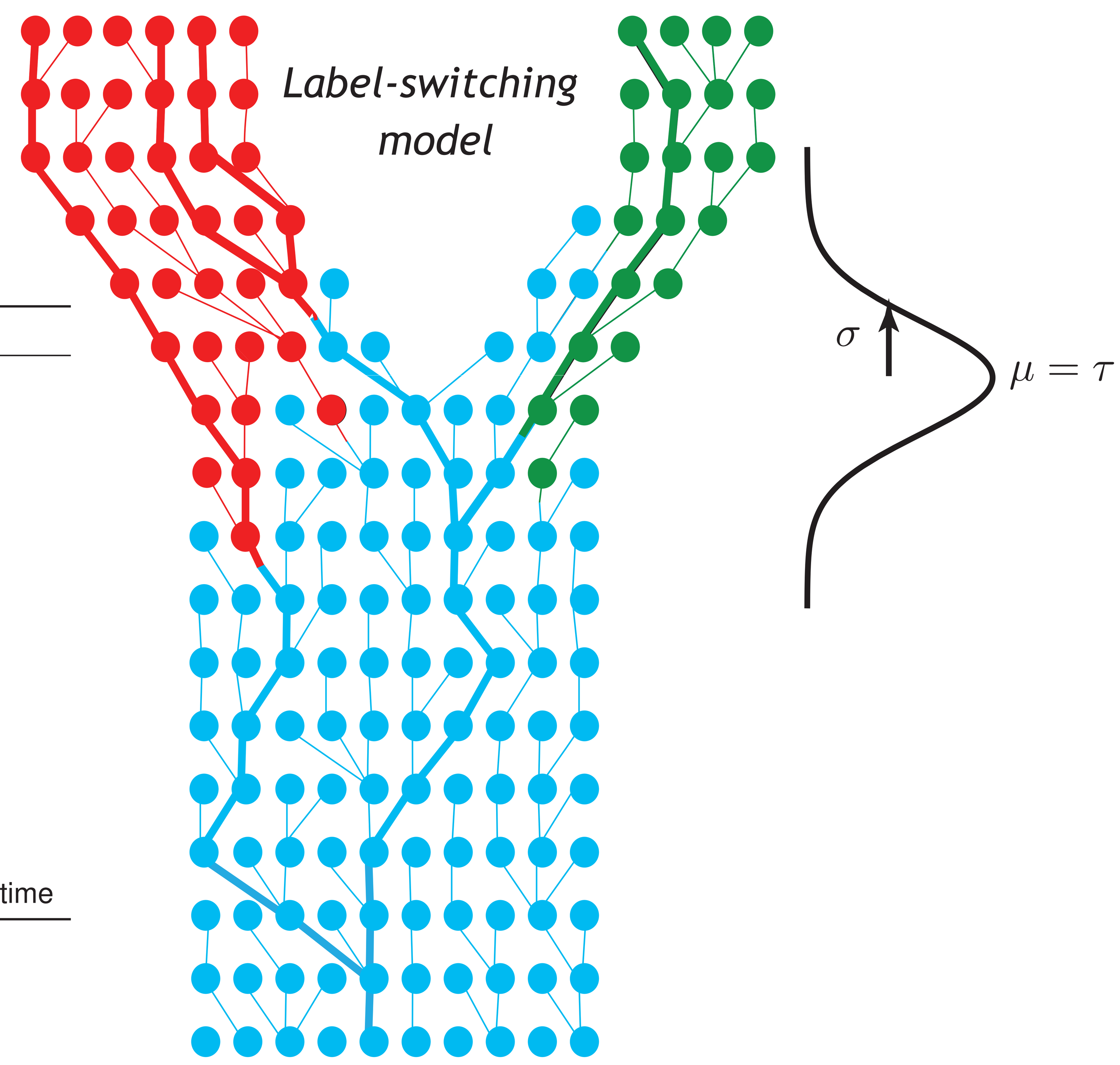
Population divergence estimation using lineage-label switching

Peter Beerli[1], Haleh Ashki[1,2], Somayeh Mashayekhi[1], Michal Palzcewski[1,3]

[1]Department of Scientific Computing, Florida State University, [2]now at RAH-foundation, [3]now at Google Inc



Parameter	Legend
Θ	Scaled Population size
N_e	Effective Population size
M	Scaled immigration rate
k	sample lineages
τ	Scaled divergence time
μ	Divergence time parameter
σ	Standard deviation
ρ	All parameters
i	Subscript for population i
λ_c	Coalescent rate
λ_m	Migration rate
λ'_s	Divergence rate
λ_s	Integrated Divergence rate
t_0	start time
u	time interval
b_1	upper bound for divergence time



In the standard model the divergence time is a fixed parameter; in inferences this parameter is changed using an arbitrary prior distribution. In contrast, the label-switching model treats the divergence time as a random varable with a truncated normal distribution with parameter μ and σ

$$\lambda_{c_i} = \frac{k_i(k_i-1)}{\Theta_i},$$

$$\lambda_{M.i} = \sum_{j=1}^{n_p} k_i M_{ji};$$

We use a hazard function to evaluate the risk of switching the label looking backwards in time; this leads to a mixture of watiting times, in the simplest case we consider coalescent and divergence events, in more complex cases we can add immigration events.

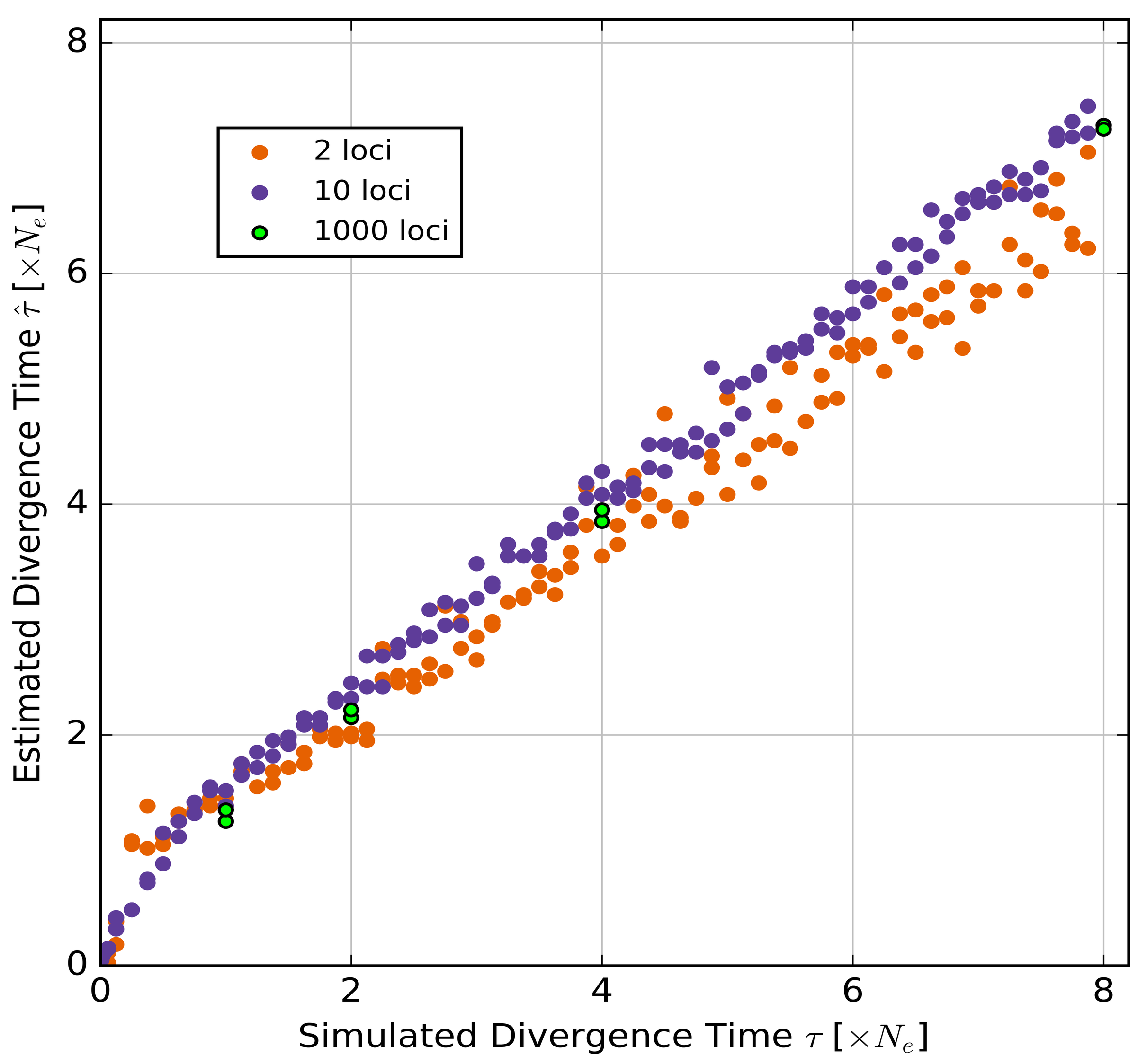
$$\lambda'_s(t) = \lambda_{\mathcal{N}(\mu,\sigma)}(t) = \frac{\sqrt{\frac{2}{\pi}}e^{-\frac{(\mu-t)^2}{2\sigma^2}}}{\sigma\left(\operatorname{erf}\left(\frac{\mu-t}{\sqrt{2}\sigma}\right)-\operatorname{erf}\left(\frac{\mu-b_1}{\sqrt{2}\sigma}\right)\right)}, \lambda_s(t_0, t_0+u) = \log\left(\frac{\operatorname{erf}\left(\frac{\mu-t_0}{\sqrt{2}\sigma}\right)-\operatorname{erf}\left(\frac{\mu-b_1}{\sqrt{2}\sigma}\right)}{\operatorname{erf}\left(\frac{\mu-(t_0+u)}{\sqrt{2}\sigma}\right)-\operatorname{erf}\left(\frac{\mu-b_1}{\sqrt{2}\sigma}\right)}\right).$$

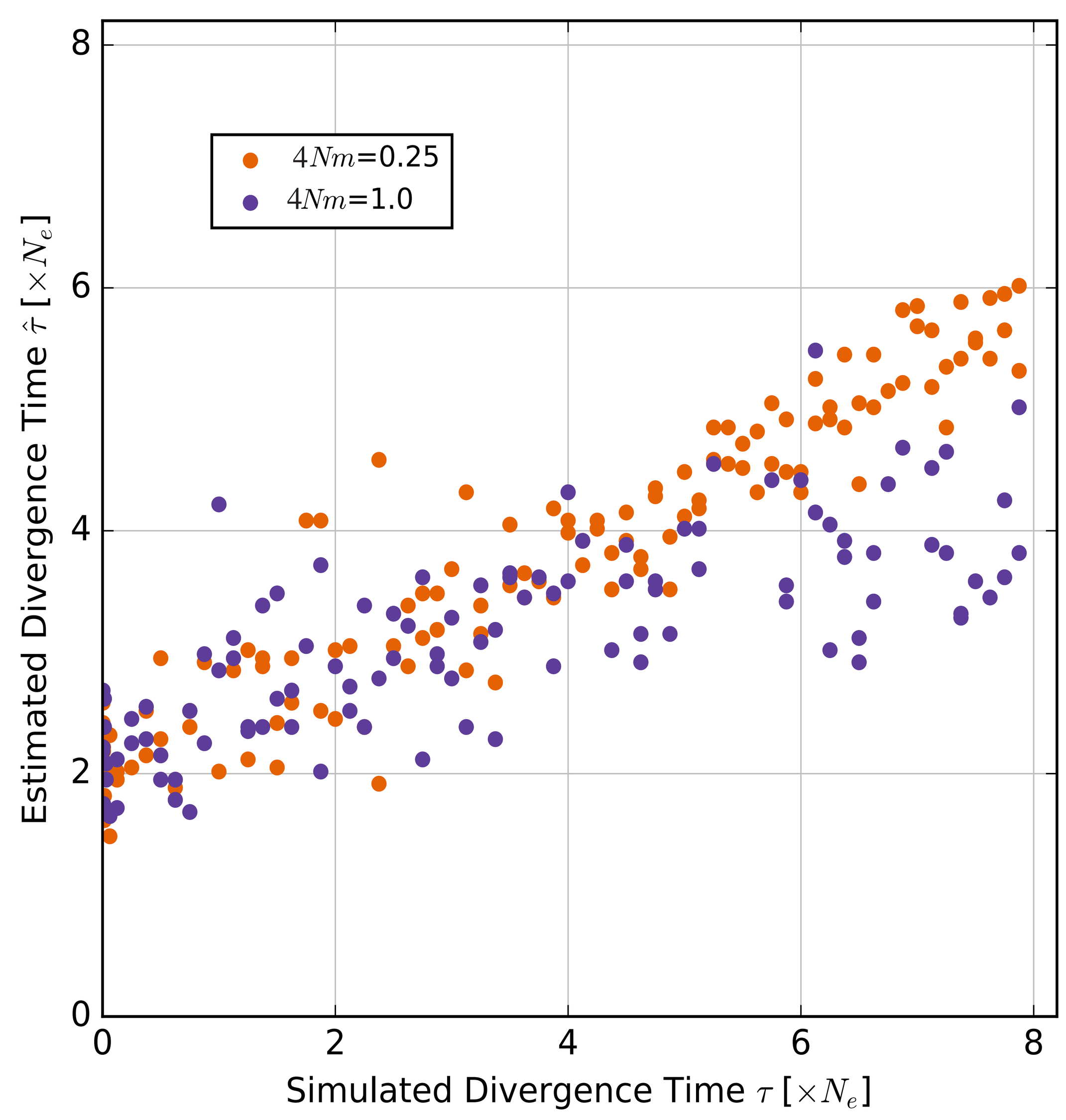
$$p_d(u|G,t_0,\rho) = \frac{1}{k}e^{-u(\lambda_c+\lambda_m)+\lambda_s(t_0+u)}\left[\lambda_c+\lambda_m+\lambda'_s(t_0+u)\right]\int_0^\infty e^{-u\lambda_c}e^{-u\lambda_m}\lambda'_s(t)\ e^{\lambda_s(t_0,t_0+u)}du.$$

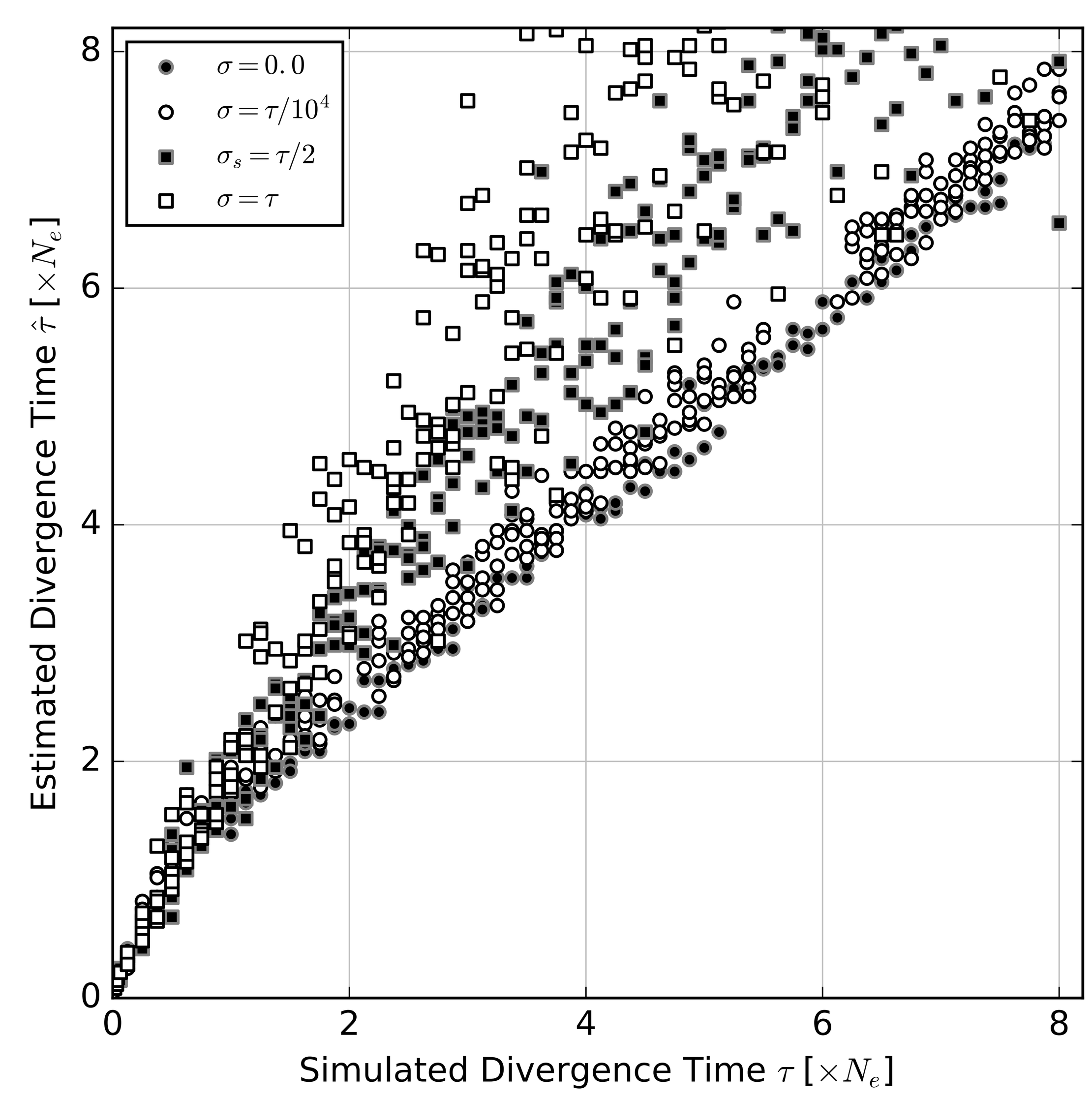
Unfortunately, the integral needs to be solved numerically. This makes the above calculation very slow. Instead of solving the integral numerically, we approximate using the midpoint rule, this approximation leads to a simpler, approximate, solution:

$$p_s(u|G,t_0,\rho) = e^{-u\left(\frac{k(k-1)}{\Theta}+kM+\lambda'_s(t_0+\epsilon)\right)}\frac{\lambda'_s(t_0+\epsilon)}{k}$$

We implemented this method in the program MIGRATE version 4, which is available from the MIGRATE website (int the preview section): http://peterbeerli.com/migrate-html5 or http://popgen.sc.fsu.edu







Disclaimer: figures will most likely change for the publication.