

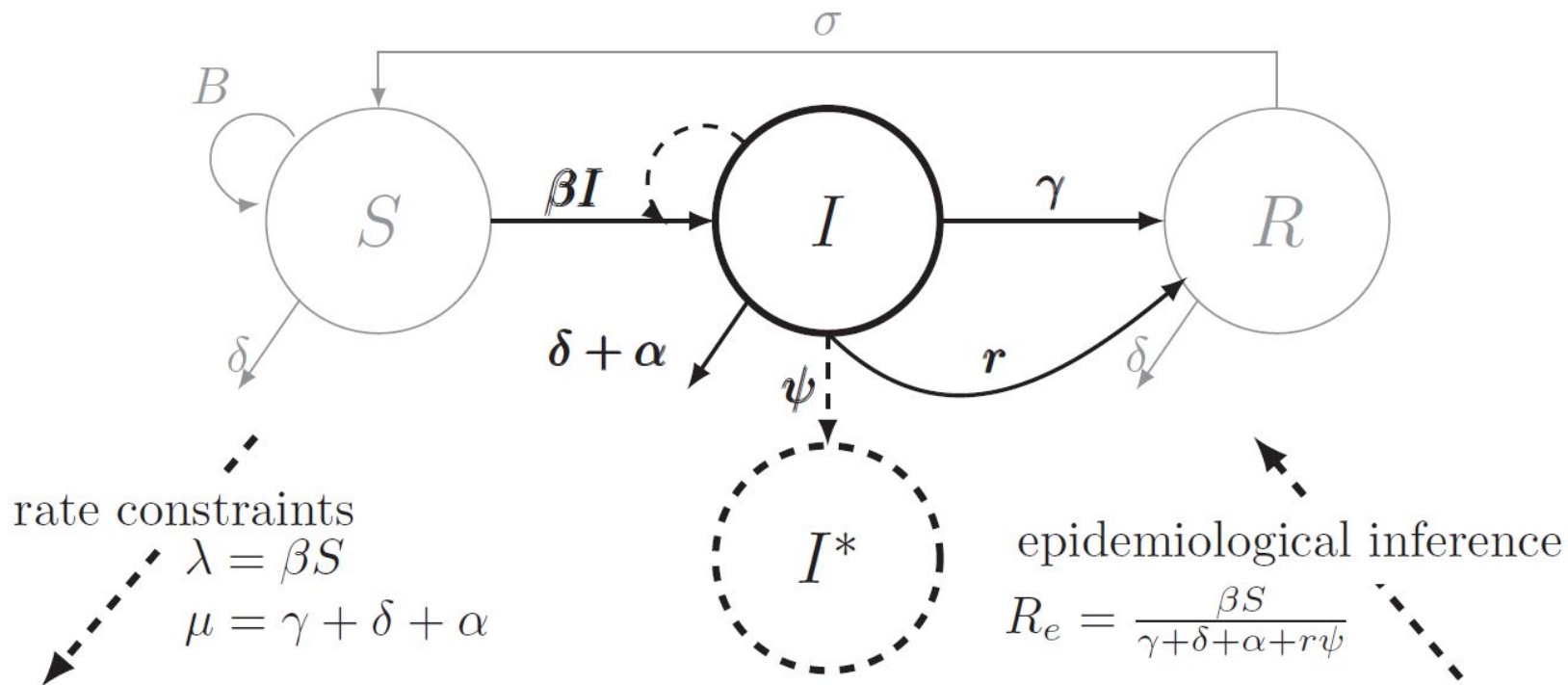
# “Unifying Phylogenetic Birth–Death Models in Epidemiology and Macroevolution” MacPherson et al (2021)

Seth D Temple  
December 10, 2021

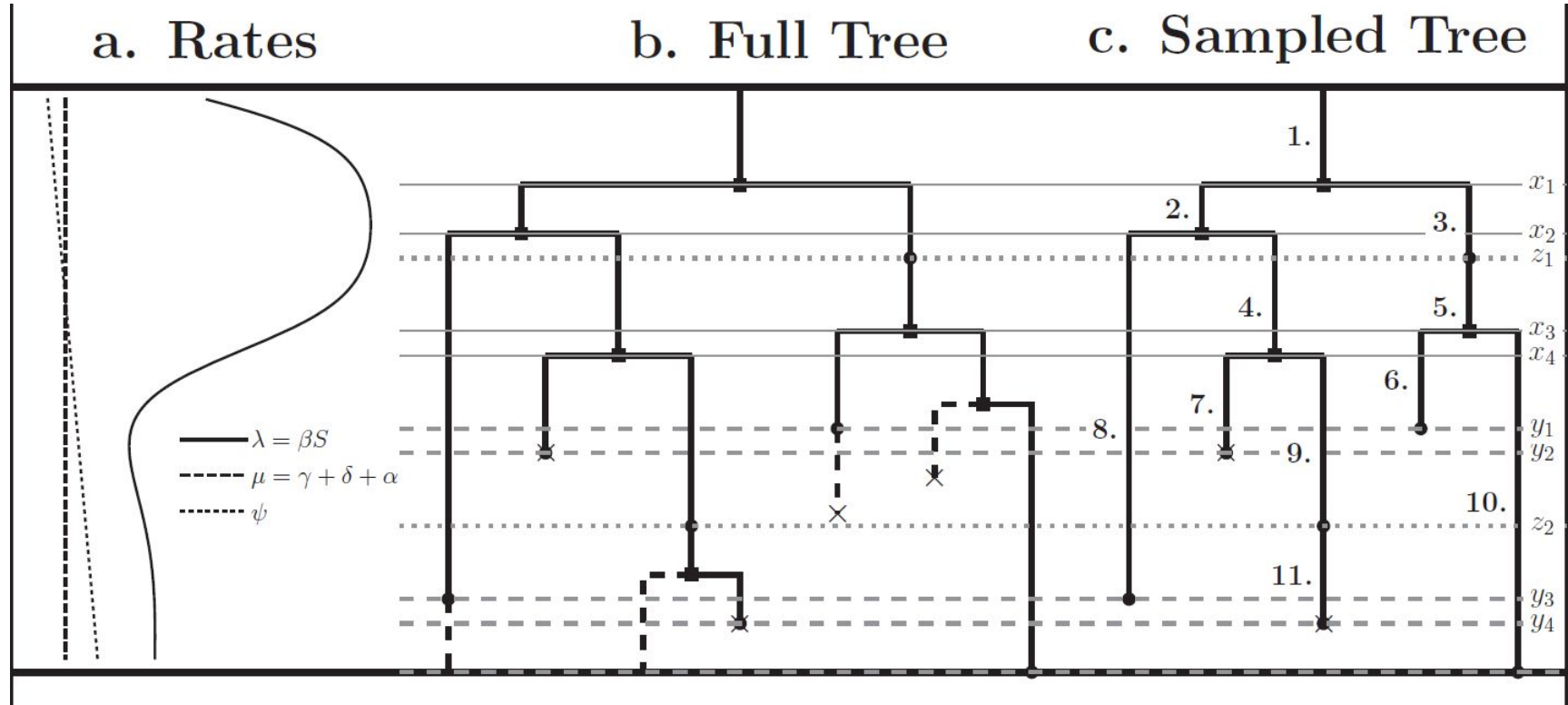
# Overview

- There are too many birth-death-sampling (BDS) models. I lose track of them!
- These makes different assumptions about parameter constraints and sampling processes.
- There is a **neat connection between** BDS models and Susceptible-Infected-Removed (SIR) models
- A **general** BDS model with a generic likelihood can cover all the BDS models common in the current literature

# SIR model



# BDS model



# Generic algorithm

1. Set up initial value problem (IVP) for prob. some edge gives rise to obs. phylo. tree between some time and present
2. Set up IVP for prob. a lineage back in time leaves no sampled descendants
3. Define the likelihood
- ~~4. Rewrite likelihood for computational application~~
- ~~5. Consider likelihoods conditional on some tree properties~~

# Step 1 in algorithm

$$\begin{aligned} g_e(\tau + \Delta\tau) \approx & \underbrace{(1 - \lambda(\tau)\Delta\tau)(1 - \mu(\tau)\Delta\tau)(1 - \psi(\tau)\Delta\tau)}_{\text{nothing happens}} \times g_e(\tau) \\ & + \underbrace{\lambda(\tau)\Delta\tau(1 - \mu(\tau)\Delta\tau)(1 - \psi(\tau)\Delta\tau)}_{\text{birth event}} \times 2g_e(\tau)E(\tau) \\ & + \underbrace{\mu(\tau)\Delta\tau(1 - \lambda(\tau)\Delta\tau)(1 - \psi(\tau)\Delta\tau)}_{\text{death event}} \times 0 \\ & + \underbrace{\psi(\tau)\Delta\tau(1 - \lambda(\tau)\Delta\tau)(1 - \mu(\tau)\Delta\tau)}_{\text{sampling event}} \times 0 + \mathcal{O}(\Delta\tau^2). \end{aligned} \tag{1}$$

# Differential equations in tree likelihood

$$\frac{dg_e(\tau)}{d\tau} = -(\lambda(\tau) + \mu(\tau) + \psi(\tau))g_e(\tau) + 2\lambda(\tau)g_e(\tau)E(\tau). \quad (3)$$

$$g_e(s_e) = \begin{cases} \lambda(s_e)g_{e1}(s_e)g_{e2}(s_e) & \text{birth event giving rise} \\ & \text{to edges } e1 \text{ and } e2 \\ (1-r(s_e))\psi(s_e)g_{e1}(s_e) & \text{ancestral sampling event} \\ \psi(s_e)r(s_e) + \psi(s_e)(1-r(s_e))E(s_e) & \text{terminal sampling event} \\ \rho_0 & s_e=0, \text{ extant sample} \end{cases}$$

# Differential equations in tree likelihood

$$g_e(\tau) = \Psi(s_e, \tau) g_e(s_e), \quad (5)$$

where the auxiliary function,  $\Psi$ , is given by:

$$\Psi(s_e, \tau) = \exp \left[ \int_{s_e}^{\tau} 2\lambda(x)E(x) - (\lambda(x) + \mu(x) + \psi(x)) dx \right]. \quad (6)$$

This function,  $\Psi(s, t)$ , maps the value of  $g_e$  at time  $s$  to its value at  $t$ , and hence is known as the probability “flow” of the Kolmogorov backward equation (Louca and Pennell



## Step 2 in algorithm

$$\begin{aligned} E(\tau + \Delta\tau) = & \underbrace{(1 - \lambda(\tau)\Delta\tau)(1 - \mu(\tau)\Delta\tau)(1 - \psi(\tau)\Delta\tau)}_{\text{nothing happens}} \times E(\tau) \\ & + \underbrace{\lambda(\tau)\Delta\tau(1 - \mu(\tau)\Delta\tau)(1 - \psi(\tau)\Delta\tau)}_{\text{birth event}} \times E(\tau)^2 \\ & + \underbrace{\mu(\tau)\Delta\tau(1 - \lambda(\tau)\Delta\tau)(1 - \psi(\tau)\Delta\tau)}_{\text{death event}} \times 1 \\ & + \underbrace{\psi(\tau)\Delta\tau(1 - \lambda(\tau)\Delta\tau)(1 - \mu(\tau)\Delta\tau)}_{\text{sampling event}} \times 0. \end{aligned} \tag{7}$$

# Tree likelihood

$$\begin{aligned}
 g_{\text{stem}}(T) = & \underbrace{\rho_0^{N_0}}_{\text{extant tips}} \underbrace{\prod_{i=1}^I \lambda(x_i)}_{\text{births}} \underbrace{\prod_{j=1}^n \left[ \psi(y_j)(1-r(y_j))E(y_j) + \psi(y_j)r(y_j) \right]}_{\text{extinct tips}} \\
 & \times \underbrace{\prod_{k=1}^m \psi(z_k)(1-r(z_k))}_{\text{ancestral samples}} \underbrace{\prod_{e \in \mathcal{T}} \Psi(s_e, t_e)}_{\text{edges}},
 \end{aligned} \tag{9}$$

Model generalizes the  
existing models in the  
current literature

# Assumptions

- All viral lineages are exchangeable
- Independence of lineages
- Piecewise continuous functions in time
- Exponentially-distributed periods, and/or most everything is a Poisson process back in time ?!?
- Diversity independence

# Questions

- How is it useful to say that BDS is an SIR model?
- How does their tree likelihood derivation compare to other derivations you are familiar with?
- Does anyone know more about this nonidentifiability issue in BDS models from extant samples only and/or serially sampled phylogenetic data?

# Questions

- Why is the reparameterization (Step 4) from edges to critical times more useful in application?
- Does anyone have a mental picture for the multi-type BDS model with anagenesis and cladogenesis?
- How well-behaved are the multivariate likelihoods?  
This is an MLE procedure.