# Introduction

This report describes the current status of the framework on developing process which aims to, as a first project, ...

The current reconstruction process

# 1 The current framework

**Input:** Reconstructed tree (observed, without species), Parameters.
**init:** Set $i = 1$
**1.** Calculate waiting time parameter for speciation $\sigma_i^s = n_i(\lambda_0 - \beta n_i)$
**2.** Draw next speciation time $t_s$ from exponential distribution with parameter $\sigma_i^s$
**if** speciation ($t_s$) ocurrs before next branching time **then**
  **3a.** Draw extinction event $t_e$ from exponential distribution with parameter $\mu_0$.
  **if** extinction occurs before current time **then**
    **4a.** Update tree with the new extincted specie
  **else**
    **4b.** Go to step 1.
  **end if**
**else**
  **3b.** Go to step 1. updating with the next branching time
**end if**
**Output:** Full tree.

## Simulations

To test this reconstruction framework we perform the following simulation experiment

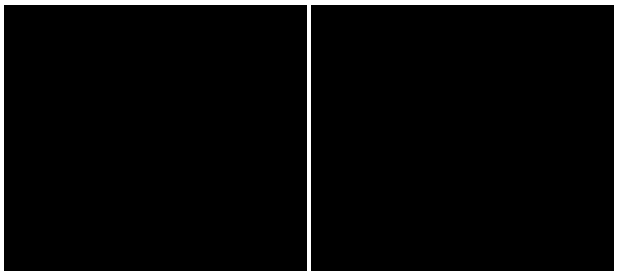
1. simulate 1000 trees and calculate their respectively MLE
2. drop extinct species for all the 1000 trees
3. for each tree reconstruct 100 trees with their respective (complete tree) MLE as parameters
4. calculate the join likelihood for each set of 100 trees

The image bellow shows the plots for each parameter. On the x-axes is the real estimation... whereas the y-axes has the estimation after the reconstruction.

**Analysis:** the reconstruction process is giving a good aproximation, however the aproximation is not enough in the sense that the new MLE is sligly biased which is an obstacle for the convergence of the EM algorithm.

The **main reason** I see about why this aproximation is not unbiased is because from the observed tree we limite the framework to...

lambda.png



## Solutions

In order to obtain a robust method with convergence on the EM algorithm I see two posible ways with limited ideas to develope them

- Continue with this aproximated algorithm described on previous section and find a way to compensate this underestimation of parameters. The problem with this option is that I do not see any way for that compensation yet. Do you have any idea?
- Propose another reconstruction model wich use all information contained on the observed tree. A draft of a posible algorithm is described bellow.

### An alternative algorithm