INPHEST Development Journal

October 29, 2015

1 Simulation of Summary Statistics

1. Generate target host tree $\bar{H}$.

   Inputs • Sequence data.

   • Divergence time calibration priors.

   Outputs • Time-calibrated phylogeny, $\bar{H}$.

   Procedure • Use `BEAST` to generate sample from the posterior distribution of trees given the data, $\mathcal{H}$.

   • Summarize $\mathcal{H}$ using MCCT or other criteria to produce $\bar{H}$.

2. Generate samples of augmented host trees, $\mathbf{h}^*$, conditioned on target host tree, $\bar{H}$.

   Inputs • Time-calibrated phylogeny, $\bar{H}$.

   • Labeling of tips on $\bar{H}$ by geography (distributional range, specified as presence/absence over set of areas, for each tip).

   Outputs • Samples of host tree augmented with biogeographical history $\mathbf{h}^*$.

   Procedure • Use `BayArea` or `RevBayes` to generate stochastic mapped samples.

3. For *each* augmented host tree, derive schedule of events, $\mathbb{E}$, consisting of ordered pairs of (τ, η), where τ is the simulation time of an event and η is the event (host speciation, host extinction, host dispersal, etc.) that occurs at that time on that particular tree.

4. For each simulation replicate:

   (a) Sample a schedule of events, $\mathbb{E}$.

(b) Simulate symbiont tree, ψ, conditioned on the schedule of events on host tree.

(c) Scale symbiont tree by each substitution rate scaling factor to yield scaled symbiont tree $\psi\prime$.

(d) Calculate summary statistics on scaled symbiont tree, $\psi\prime$.