Simulating, inferring\(^+\), and analysing time-calibrated phylogenies

\(^+\)inferring: see talk yesterday
Estimating macro-evolutionary processes

Major challenge in evolutionary biology, conservation biology & paleontology:

- Identify main factors governing macro-evolutionary dynamics
Estimating macro-evolutionary processes

Major challenge in evolutionary biology, conservation biology & paleontology:

- Identify main factors governing macro-evolutionary dynamics

Phylogenetic tree of apes
(from Yang & Rannala, MBE, 2005)
Macro-evolutionary models

- Constant speciation and extinction rates
Macro-evolutionary models

- Constant speciation and extinction rates

Null model

Rate changes due to extrinsic factors

environmental-dependence
(Stadler, PNAS, 2011; Morlon et al., PNAS, 2011)

density-dependence

http://scientopia.org/blogs/ethicsandscience/
Macro-evolutionary models

- Constant speciation and extinction rates

Null model

Rate changes due to extrinsic factors

- environmental-dependence
  (Stadler, PNAS, 2011; Morlon et al., PNAS, 2011)
- density-dependence

Rate changes due to intrinsic factors

- trait-dependence

http://scientopia.org/blogs/ethicsandscience/
Macro-evolutionary models

- Constant speciation and extinction rates

Null model

Rate changes due to extrinsic factors

- environmental-dependence
  (Stadler, PNAS, 2011; Morlon et al., PNAS, 2011)

- density-dependence

Rate changes due to intrinsic factors

- trait-dependence
Maximum likelihood approach for inferring macro-evolutionary processes

Mathematical
- Define a macro-evolutionary model
- Find efficient way to calculate likelihood function: Probability of the phylogenetic tree given the parameters

Computational
- Implement R packages TreePar & TreeSim
- Parameter inference on cluster

Empirical

Mammals

Birds

Stadler (PNAS, 2011); Stadler (Syst. Biol., 2011)
The birth-death model as a model for speciation and extinction

- stem age of a clade $t$
- speciation rate $\lambda$
- extinction rate $\mu$
- sampling probability $\rho$

Rates may depend on:
1) time (environmental-dependence)
2) # of species (density-dependence)

Stadler (JTB, 2009)
The birth-death model as a model for speciation and extinction

- stem age of a clade $t$
- speciation rate $\lambda$
- extinction rate $\mu$
- sampling probability $\rho$

Rates may depend on:

1) time (environmental-dependence)
2) # of species (density-dependence)

Stadler (JTB, 2009)
1) Modeling environmental effects

Stadler (PNAS, 2011)
1) Modeling environmental effects

Calculate $\text{Prob (tree | parameters)}$ of reconstructed phylogeny by “integrating” over all complete phylogenies:

Efficiently done using differential equations and recursions

Stadler (PNAS, 2011)
Likelihood formula

\[
\text{Lik}(\mathcal{X}) = \rho^n (\lambda_0 - \mu_0)^{2n} \frac{q_l(x_1)(x_1)^2}{(1 - p_l(x_1)(x_1))^2} \prod_{i=2}^{n-1} \lambda_l(x_i)q_l(x_i)(x_i) \prod_{i=1}^{m} (\lambda_i - \mu_i)^{2n} q_{i-1}(t_i)^{n_i}
\]

...with the \( x_i \) being the branching times and \( p(x_i), q(x_i), l(x_i) \)

being known functions depending on the parameters \( \lambda, \mu, \rho, t \)
Likelihood formula

\[
\text{Lik} \left( \begin{array}{c}
\vdots \\
\vdots \\
\vdots \\
\vdots \\
\vdots \\
\end{array} \right) = \rho^n (\lambda_0 - \mu_0)^{2n} \frac{q_l(x_1)(x_1)^2}{(1 - p_l(x_1)(x_1))^2} \prod_{i=2}^{n-1} \lambda_l(x_i)q_l(x_i)(x_i) \prod_{i=1}^{m} (\lambda_i - \mu_i)^{2n_i} q_{l-1}(t_i)^{n_i}
\]

Likelihood function used for maximum likelihood inference

...with the \( x_i \) being the branching times and \( p(x_i), q(x_i), l(x_i) \)

being known functions depending on the parameters \( \lambda, \mu, \rho, t \)
The maximum likelihood parameters for a species tree are estimated for $m = 0, 1, 2...$ rate shifts.

The most likely number of shifts is determined using likelihood ratio tests.

Method implemented in R library TreePar (available on CRAN).
Analyses in R

- TreeSim
- TreePar
Species phylogeny rejects increased mammalian diversification at KT-boundary

Mammalian phylogeny

(from Bininda et al., Nature, 2007)

Stadler (PNAS, 2011); Meredith,...,Stadler et al. (Science, 2011)
Species phylogeny rejects increased mammalian diversification at KT-boundary

(diversification = speciation $\lambda$ - extinction $\mu$)

Mammalian phylogeny
(from Bininda et al., Nature, 2007)

Null model: constant rates bad fit
Environmental model: 4 shifts ($p=1.0$)
Simulation results

Stadler (PNAS, 2011); Meredith,...,Stadler et al. (Science, 2011)
2) Modeling density-dependent effects

Model

N: # of species
K: carrying capacity
\( \lambda(N) = \lambda_{\text{const}} \left(1 - \frac{N}{K}\right) \)

Etienne, Haegeman, Stadler et al. (Proc. Roy. Soc. B, 2012); Leventhal, ..., Stadler (MBE, 2014)
2) Modeling density-dependent effects

Model

\[ \lambda(N) = \lambda_{\text{const}} (1-N/K) \]

Likelihood

\[
\text{Lik} \left( \begin{array}{c}
\text{\ldots}
\end{array} \right) = \prod_{i=2}^{n} \left( L e^{M_i(x_{i-1} - x_i)} \right) e_n
\]

with \( L \) and \( M_i \) matrices of dimension \( K \), \( e_n \) vector.

Etienne, Haegeman, Stadler et al. (Proc. Roy. Soc. B, 2012); Leventhal, ..., Stadler (MBE, 2014)
2) Modeling density-dependent effects

- **N:** # of species
- **K:** carrying capacity
- \( \lambda(N) = \lambda_{\text{const}} (1 - N/K) \)

\[
\text{Lik} \left( \begin{array}{c}
\vdots \\
\end{array} \right) = \prod_{i=2}^{n} \left( Le^{M_i(x_{i-1} - x_i)} \right) e_n
\]

- with L and \( M_i \) matrices of dimension K, \( e_n \) vector.

**Evaluating vectors applied to matrix exponentials:**
- Method by Al-Mohy & Higham (2011)

Etienne, Haegeman, Stadler et al. (Proc. Roy. Soc. B, 2012); Leventhal, ..., Stadler (MBE, 2014)
Incomplete sampling schemes

Stadler (PNAS, 2011); Stadler & Bokma (Syst. Biol., 2013)
Incomplete sampling schemes

So far:
Random sampling

Stadler (PNAS, 2011); Stadler & Bokma (Syst. Biol., 2013)
Incomplete sampling schemes

So far:
Random sampling

Often more realistic:
Biased sampling

Stadler (PNAS, 2011); Stadler & Bokma (Syst. Biol., 2013)
Incomplete sampling schemes

So far:
Random sampling

Often more realistic:
Biased sampling

Stadler (PNAS, 2011); Stadler & Bokma (Syst. Biol., 2013)
Questions?