Lineage diversification models

Ricardo Betancur
(some slides courtesy of M. Alfaro & L. Revell)
A fundamental pattern of biodiversity: unvenness
Ecological theory on adaptive radiations

• Simpsonian adaptive radiations often occur when a lineage encounters ecological opportunity in the form of:
  • Colonization of novel habitats
  • Extinction of competitors
  • Evolution of intrinsic traits (key innovations)

• Adaptive radiations often result in burst of diversification as lineages entering novel adaptive zones become decoupled from the normal diversity-dependent controls that regulate rates
Diversity-dependent diversification & adaptive radiation

African cichlids

Hawaiian honeycreepers

++ speciation
++ phenotypic evolution
Outline

- Pure-birth model: lineage through time plots, $\gamma$ statistic and lineage diversification index
- Constant birth-death model: estimating speciation and extinction rates from molecular phylogenies
- Complex models: clade-dependent, time-dependent (incl. diversity-dependent) & trait-dependent models
- State dependent diversification
Outline

- Pure-birth model: lineage through time plots, \( \gamma \) statistic and lineage diversification index
Lineage diversification: basic terminology

Speciation or cladogenesis or origination: lineage birth

Extinction: lineage death

Net diversification: speciation (\(\lambda\)) - extinction (\(\mu\))

Lineage turnover: extinction (\(\mu\)) / speciation (\(\lambda\))
Yule process or pure birth (pb)

- Assumes constant rate of speciation across lineages
- No extinction
Under a pure-birth model of species accumulation, the number of lineages accrues log-linearly with time.
Lineage diversification: lineage through time plots
Lineage accumulation under pure-birth

Under a pure-birth model of species accumulation, the number of lineages accrues log-linearly with time.

Furthermore, the slope of this line has an expected expectation that is equal to the speciation rate.
Diversification rate analysis: The $\gamma$ statistic

A simple statistic was devised by Pybus & Harvey (2000) and it is called the gamma ($\gamma$) statistic.

Gamma is calculated as follows:

$$
\gamma = \frac{1}{n-2} \sum_{i=2}^{n-1} \sum_{k=2}^{i} kg_k - \left( \frac{T}{2} \right), \quad T = \sum_{j=2}^{n} jg_j
$$

$g$ is just a vector containing the internode distances (absolute or relative ages) from our phylogeny.

Though complicated, this statistic is useful because it has a standard normal distribution for pure-birth trees.
Diversification rate analysis: The $\gamma$ statistic

$\gamma$ is just a vector containing the internode distances (absolute or relative ages) from our phylogeny. Also called waiting times.
Diversification rate analysis: The $\gamma$ statistic

- Measures the relative position of nodes within a phylogeny
- Function of $n$ nodes and $g$ (summed branch lengths within a given time interval)
Interpreting the $\gamma$ statistic

Gamma of 0 indicates that nodes have a balanced distribution, relative to PB expectation.

Significantly (-) gamma suggests that the speciation rate was initially high, but then slowed through time.

Significant (+) gamma suggests the opposite; i.e., that the speciation rate was initially slow, and increased over time.

Since the statistic has a standard normal distribution we can test hypotheses about gamma (e.g., using a Z-test).
Interpreting the $\gamma$ statistic
Interpreting the $\gamma$ statistic

Gamma 0  Gamma (-)  Gamma (-)**
Interpreting the $\gamma$ statistic

One caveat that needs to be noted is that positive gamma can also be caused by unseen past extinction. This is due to a phenomenon that is known as the “pull of the present” (discussed later).

Incompletely sampled phylogenies can also artificially affect gamma estimates to be negative.
Example of the γ statistic: marine-derived freshwater ariids
$F \rightarrow M = 1 \quad M \rightarrow F = 13$

- Ariids
- Other catfishes

- Red: Marine
- Blue: Freshwater
- Purple: Diadromous

Ma: 62.5, 50, 37.5, 25.0, 12.5, 0
Colonization of novel habitats promote adaptive radiation (Simpson, 1953)
Is ecological opportunity presented upon colonization of novel ecological regimes?

- Mesoamerica: 6 vs. 1100
- South America: 11 vs. 6600
- Africa: 2 vs. 3600
- South East Asia: 15 vs. 3600
- Australia-New Guinea: 28 vs. 600

Freshwater ariids
Remaining freshwater fishes
Unevenness in the diversity of freshwater ariid clades
Example of the $\gamma$ statistic: marine-derived freshwater ariids in Australia & New Guinea

Example of the $\gamma$ statistic: marine-derived freshwater ariids in Australia & New Guinea
Example of the $\gamma$ statistic: marine-derived freshwater ariids in Australia & New Guinea

What about incomplete sampling?

With random sampling of four taxa, what nodes would be expected to miss?

How would that affect gamma?
MCCR: the Monte Carlo Constant Rates test for incompletely sampled trees

Incomplete sampling: old nodes are more likely to be sampled, biasing gamma towards negative values

Solution MCCR test:

- Simulate trees with $n$ taxa
- Randomly prune missing spp.
- Calculate gamma on pruned tree
- Repeat
- Calculate corrected $P$ value

Example of the $\gamma$ statistic: marine-derived freshwater ariids in Australia & New Guinea

(a) Ariidae – complete tree
\textbf{Gamma: 1.93, $P = 0.026$}

(b) AU-NG clade
\textbf{Gamma: -3.1, $P = 0.001$}

\textit{Betancur-R. et al. (2012) - Ecol. Let.}
The lineage diversification index (LDI)

Similar to the gamma statistic

Estimates whether lineage accumulation departs from a pure-birth by assessing the difference between the LTT curve of a phylogeny and the LTT curve of a pure-birth model

(+): value: lineage experienced greater than expected diversification rates early in its history

(-): value: lineage experienced greater than expected diversification rates late in its history

Harmon et al. (2003) - Science
The lineage diversification index (LDI)

**Part (a) Ariidae – complete tree**
LDI: -0.48

**Part (b) AU-NG clade**
LDI: 0.70

*Betancur-R. et al. (2012) - Eco. Lett*
Issues with γ-statistic and LTT plots

- Branch-length estimation biases
  Model misspecification (Revell, 2005)

- Non-random species sampling
  (Cusimano & Renner, 2010)

- Time-dependent rate variation
  Violation of a constant-rate model through time (Yule, constant bd)

- Among-lineage rate variation
  Violation of a constant-rate model across lineages (Yule, constant bd)
Yule process or pure birth (pb)

- Assumes constant rate of speciation across lineages
- No extinction
Lineage accumulation under pure-birth

Under a pure-birth model of species accumulation, the number of lineages accrues log-linearly with time.

Furthermore, the slope of this line has an expectation that is equal to the speciation rate.
Yule process or pure birth (pb): estimating the birth rate

- Given clade age and species richness, the birth rate can be estimated using the Method-of-Moments equation by Magallon & Sanderson (2001):

\[ \hat{r}_{crown} = \frac{\log(n) - \log(2)}{\text{time}} \]

- Expected diversity under pb:

\[ N(t) = N(o)e^{bt} \]
Estimating the birth rate: example using *Dianthus*

- 200 spp.
- 0.9–2.1 Myr
- 2.9–7.6 spp. / Myr

\[ N(t) = N(o)e^{bt} \]