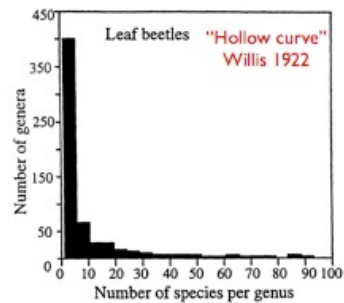


Birth-death Models and Tree Balance

Lecture 2
March 12, 2009



David Raup



Stephen J. Gould

Thomas Schopf

The "Woods Hole Group"



Outline



- Review of BD model
- Diversity of sister clades
- Tree balance
- Newer approaches



Outline



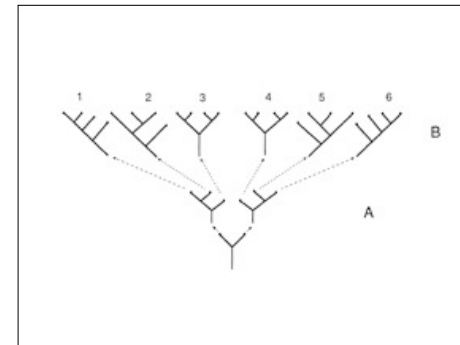
- Review of BD model
- Diversity of sister clades
- Tree balance
- Newer approaches

Birth-death model

- Assumptions:
 - Speciation and extinction occur randomly
 - Each lineage has an equal and constant rate of speciation, b , and extinction, d

ERM Model

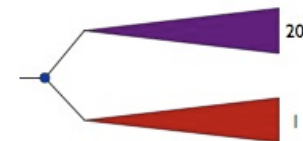
- BD is a special case of the Equal Rates Markov model (ERM)
- ERM Model assumptions
 - Markov model: next step depends only on current state
 - Equal rates: b and d are constant across lineages at any given time
 - b and/or d might change through time



★ Outline ★

- Review of BD model
- Diversity of sister clades
- Tree balance
- Newer approaches

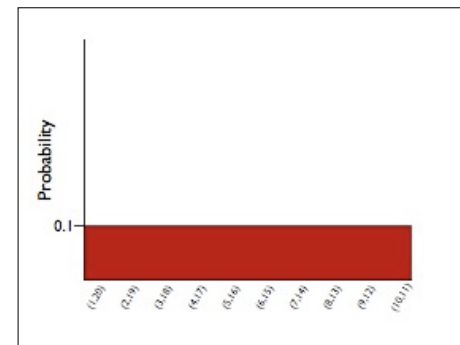
Balance of a node



Diversity of Sister Clades



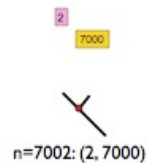
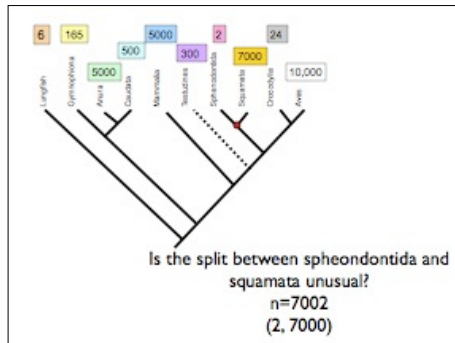
What is the expected distribution for n_1 and n_2 under BD?



Probability of a split at one node

$$\Pr[(k, n-k)] = \begin{cases} \frac{2}{n-1} & k \neq n-k \\ \frac{1}{n-1} & k = n-k \end{cases}$$

* Under any ERM model, including BD



What is the probability of obtaining this split, or something even more extreme, under the ERM model?

$$P = \Pr[(2, 7000)] + \Pr[(1, 7001)]$$

$$P = 2/(7002-1) + 2/(7002-1)$$

$$P = 0.00057$$

The Slowinski-Guyer Test

- Assumes ERM model for diversification
- Requires strikingly large differences between clades for significance
- If applied haphazardly, danger of multiple comparisons



Outline



- Review of BD model
- Diversity of sister clades
- Tree balance
- Newer approaches

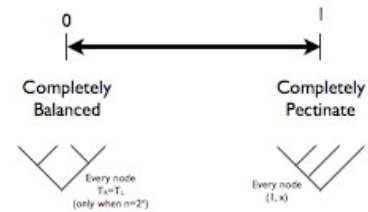
Tree Balance

- How do we summarize balance across a whole phylogenetic tree?
- How do we calculate the "null" expectation for this statistic?
- How do we compare this distribution to real data?

Tree Balance

- How do we summarize balance across a whole phylogenetic tree?
- How do we calculate the "null" expectation for this statistic?
- How do we compare this distribution to real data?

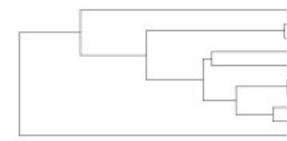
I_c (Colless 1982)



Tree Balance

- What is the "average imbalance" of all nodes in a tree?
- There are many ways to calculate this
- I will introduce one of the first, but be aware that there are different measures with different statistical properties

Example



I_c (Colless 1982)

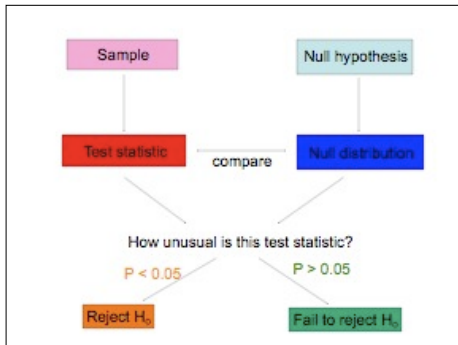
$$I_c = \frac{\sum_{\text{(all interior nodes)}} |T_R - T_L|}{\frac{(n-1)(n-2)}{2}}$$

T_R, T_L = number of taxa subtended by "left" and "right" branches

Maximum possible value for I_c with a completely pectinate tree

Tree Balance

- How do we summarize balance across a whole phylogenetic tree?
- How do we calculate the "null" expectation for this statistic?
- How do we compare this distribution to real data?



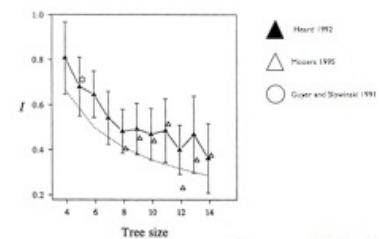
Tree Balance

- How do we summarize balance across a whole phylogenetic tree?
- How do we calculate the "null" expectation for this statistic?
- How do we compare this distribution to real data?

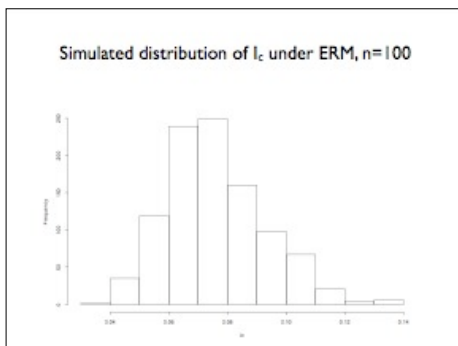
$$E[l_c] = \begin{cases} \frac{2n}{(n-1)(n-2)} \sum_{j=2}^{n/2} \frac{1}{j} & (n \text{ even}) \\ \frac{2n}{(n-1)(n-2)} \left[\frac{1}{n} + \sum_{j=2}^{(n-1)/2} \frac{1}{j} \right] & (n \text{ odd}) \end{cases}$$

(Heard 1992)

Balance of Real Trees



(Mooers and Heard 1997)



Possible Explanations

- Methodological artifact
 - Sampling
 - Phylogenetic reconstruction methods
- Variation in speciation and/or extinction across lineages



Outline



- Review of BD model
- Diversity of sister clades
- Tree balance
- Newer approaches

Sister-clade Comparisons



Does a given trait influence diversification rates?

Sister-clade Comparisons

- Compare pairs of sister clades, where one clade has the trait of interest and the other does not
- Under ERM, either clade is equally likely to be the larger of the two
- Use binomial test to evaluate significance

Binomial Probability

$$Pr(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

x = number of "successes"
n = number of trials (sister pairs)
p = 0.5 (Pr[success] under H₀)

Example



actinomorphic



zygomorphic

Example

- Zygomorphic group more diverse in 15 of 19 comparisons

$$Pr(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

$$Pr(x=15) = \binom{19}{15} 0.5^{15} 0.5^4$$

$$Pr(x=15) = 3876 * 0.5^{15} 0.5^4 = 0.007$$

$$Pr(x \geq 15) = 0.019$$

More developments

- Different tree balance statistics: N , σ^2_N , B_1 , B_2 , R , etc.
- Methods to identify particularly imbalanced nodes in a phylogenetic tree: SymmeTREE (Chan and Moore 2002)
- Other aspects of tree shape: number of cherries, etc.
- More analytic results