Birth-death Models and Tree Balance

Lecture 2
March 12, 2009

Leaf beetles "Hollow curve" Willis 1922

Number of genera vs. Number of species per genus
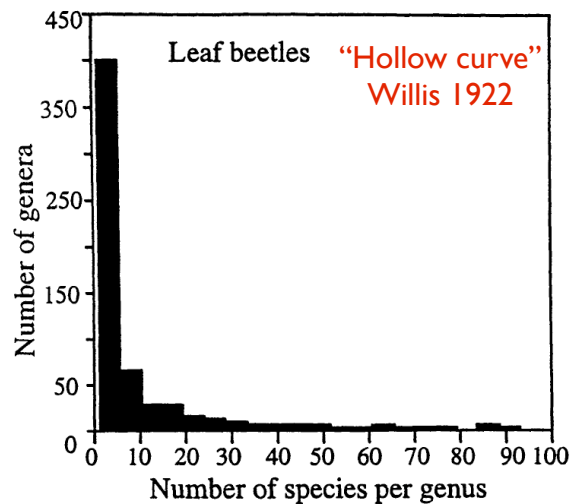
David Raup
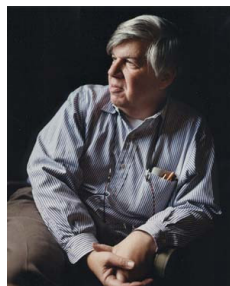
Stephen J. Gould

Thomas Schopf
Dan Simberloff

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
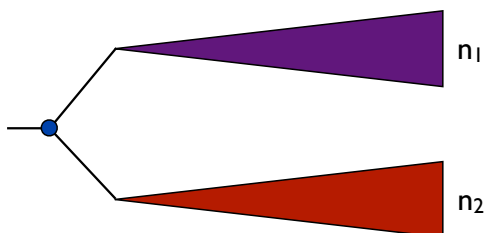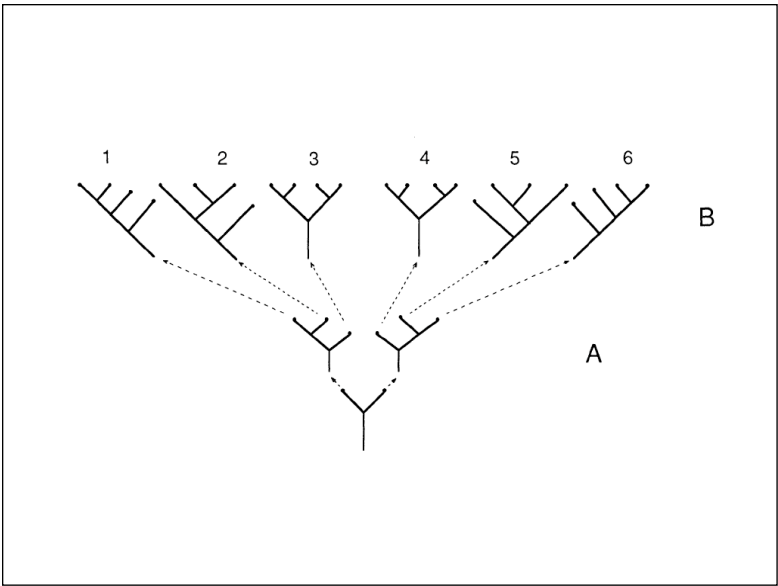
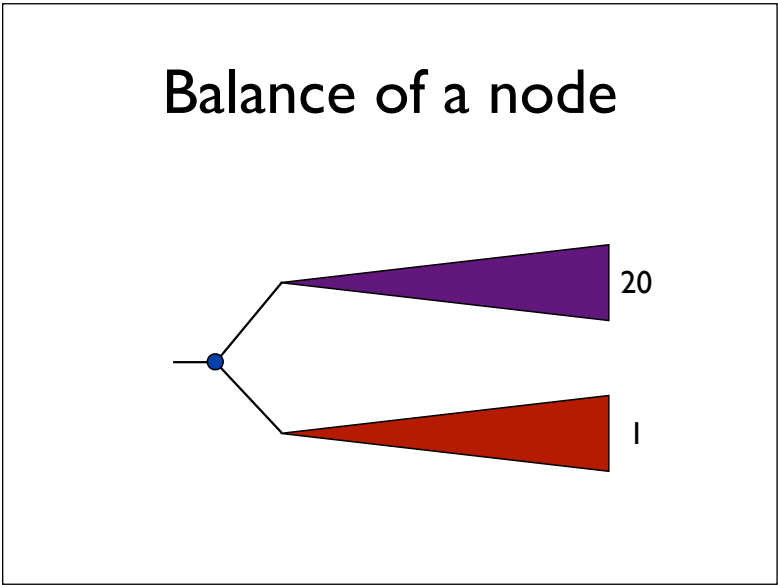 - b and/or d might change through time

Outline

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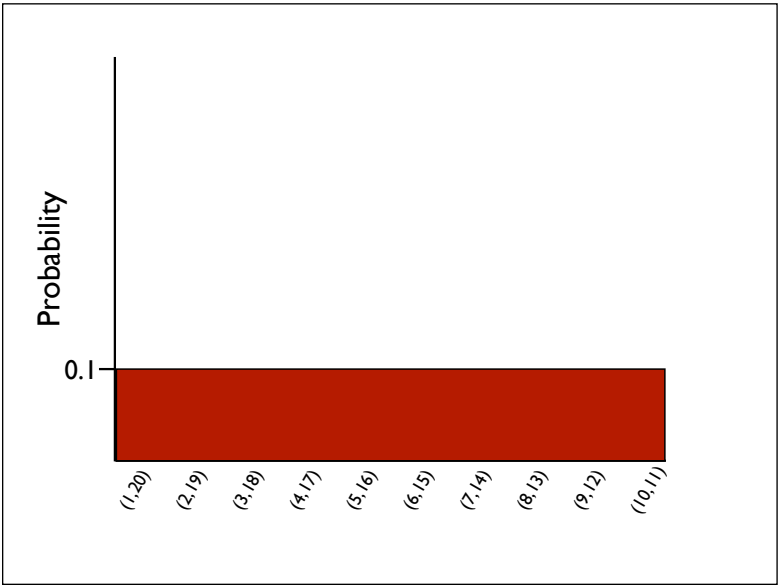
Diversity of Sister Clades

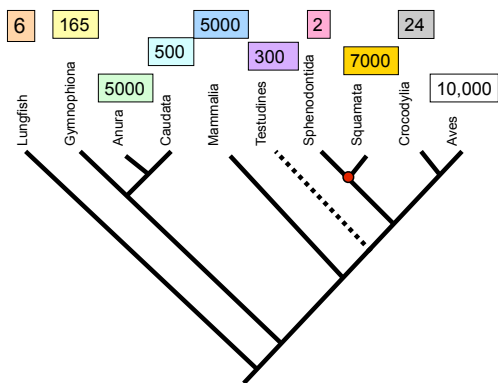What is the expected distribution for n_1 and n_2 under BD?

1 2 3 4 5 6

B

A

Balance of a node

20

I

Probability

0.1

(1,20) (2,19) (3,18) (4,17) (5,16) (6,15) (7,14) (8,13) (9,12) (10,11)

Probability of a split at one node

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

* Under any ERM model, including BD

---

6 165 5000 2 24

500 300 7000

5000 10,000

Lungfish Gymnophiona Anura Caudata Mammalia Testudines Sphenodontida Squamata Crocodylia Aves

Is the split between spheondontida and squamata unusual?
n=7002
(2, 7000)

2

7000

n=7002: (2, 7000)

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?

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
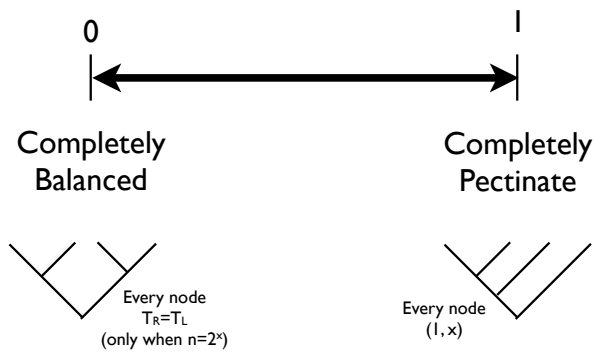
I_c (Colless 1982)

$$I_c = \frac{\sum\limits_{\text{(all interior nodes)}} |\, T_R - T_L \,|}{\dfrac{(n-1)(n-2)}{2}} ;$$
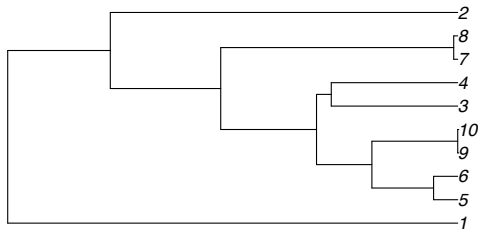
T_R, T_L = number of taxa subtended by "left" and "right" branches

Maximum possible value for I_c with a completely pectinate tree

I$_c$ (Colless 1982)

0   1

Completely Balanced   Completely Pectinate

Every node
$T_R = T_L$
(only when n=2^x)

Every node
(1, x)

Example

2
8
7
4
3
10
9
6
5
1

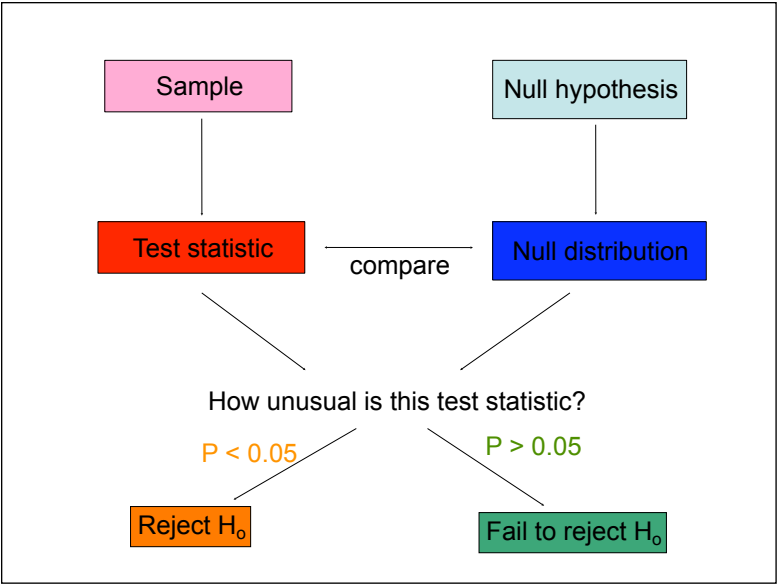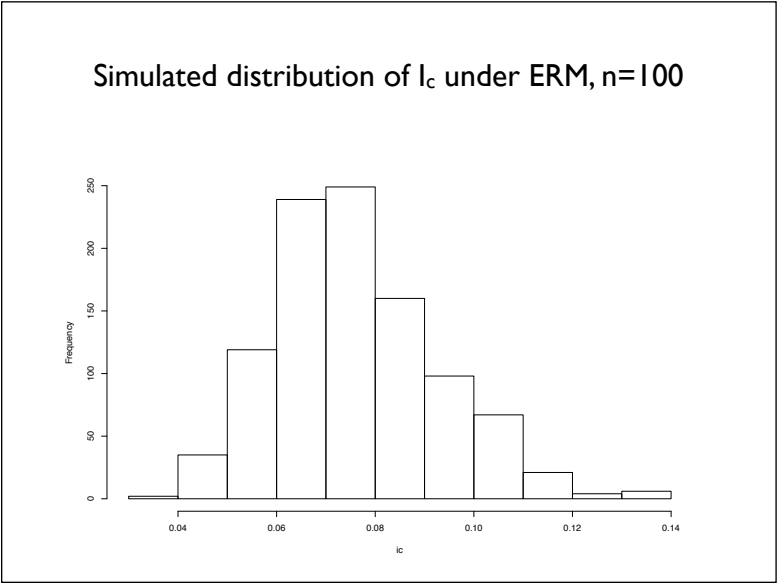
Tree Balance

- 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?

Slide 1

```
     Sample            Null hypothesis


  Test statistic  —compare—  Null distribution


        How unusual is this test statistic?

   P < 0.05                    P > 0.05

   Reject H_o                  Fail to reject H_o
```

Slide 2

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

(Heard 1992)

Slide 3

Simulated distribution of I_c under ERM, n=100
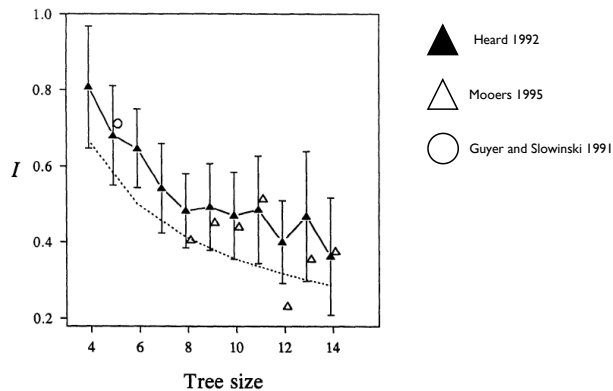
Tree Balance

- 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?

Balance of Real Trees

(Mooers and Heard 1997)

Possible Explanations

- Methodological artifact

  - Sampling

  - Phylogenetic reconstruction methods

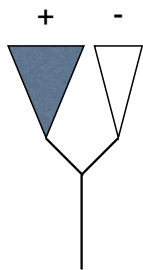- Variation in speciation and/or extinction across lineages

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
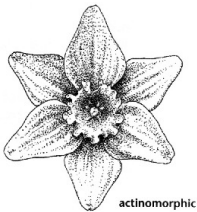- Use binomial test to evaluate significance

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 H_o)

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