

Overview of Comparative Methods

- Learning from the trees
- General statistical framework
- Basic models of growing trees and changing characters
- Details of Birth-Death



General statistical framework

- We will focus on a model fitting approach
- Develop mathematical models
- Calculate likelihoods
- Compare the fit of a set of competing models

Core Models

- Birth-death model for diversification
- Mk model for discrete characters
- Brownian motion model for continuous characters

Markov Models

- Most of the models used in comparative biology satisfy the Markov property:
- The conditional probability distribution of future states of the process, given the present state and a constant number of past states, depends only upon the present state and the given states in the past, but not on any other past states

Expected species diversity under a birth-death model:

$$E[N_t] = N_0 e^{(b-d)t}$$

N_t = species diversity after time t

N_0 = starting species diversity

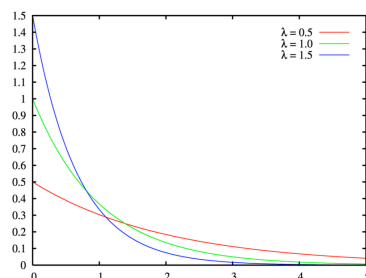
b = birth rate

d = death rate

t = time

Waiting times

- Under a birth-death model, the time intervals between successive events are always drawn from exponential distributions



Growing birth-death trees

- Same protocol, except waiting times are drawn from exponential distribution with $\lambda = n(b + d)$, where n is the number of living species at that time
- Choose a random lineage
- Decide whether you have a speciation or extinction event
- $\text{Pr}[\text{speciation}] = b/(b+d)$; $\text{Pr}[\text{extinction}] = d/(b+d)$



Outline



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

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

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

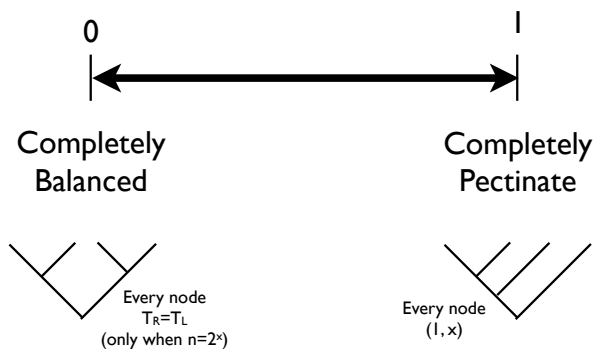
I_c (Colless 1982)

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

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

Maximum possible value for I_c with
a completely pectinate tree

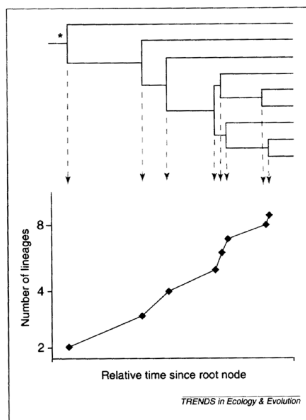
I_c (Colless 1982)



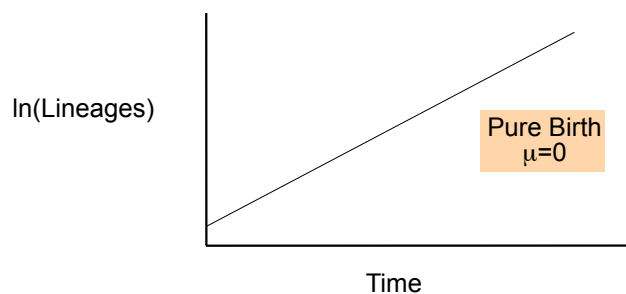
Outline

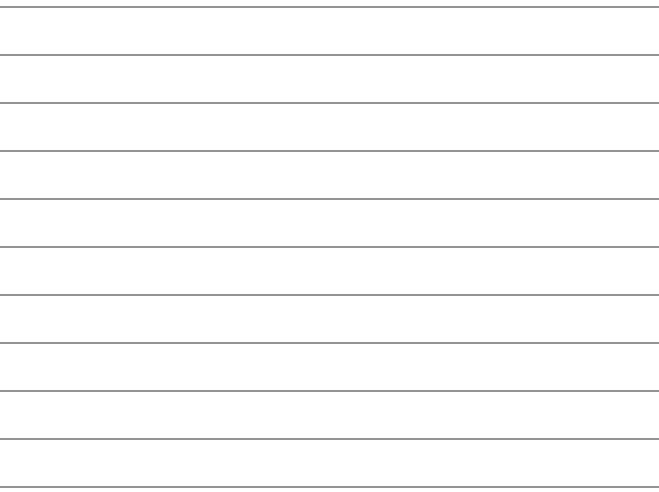
- How do lineages accumulate under the birth-death model?
- Estimating speciation and extinction rates from trees
- Testing hypotheses about diversification

Lineage-through-time



LTT Plot





This image shows a blank sheet of white paper with horizontal ruling lines. The lines are evenly spaced and run across the width of the page. There are no margins, text, or other markings on the paper.

[illegible]

ΔAIC	Interpretation
0-3	no support
3-7	weak
7-10	intermediate
>10	strong

Slowdown statistic

- Tests for longer branches near the tips of a phylogenetic tree
- Signature of an early, rapid radiation

$$\gamma = \frac{\left(\frac{1}{n-2} \sum_{i=2}^{n-1} \left(\sum_{k=2}^i k g_k \right) \right) - \left(\frac{T}{2} \right)}{T \sqrt{\frac{1}{12(n-2)}}}, T = \left(\sum_{j=2}^n j g_j \right).$$

Slowdown statistic

- Under pure-birth model, $\gamma \sim N(0,1)$
- $\gamma < 0$ means diversification has slowed
- $\gamma > 0$ difficult to interpret

$$\gamma = \frac{\left(\frac{1}{n-2} \sum_{i=2}^{n-1} \left(\sum_{k=2}^i k g_k \right) \right) - \left(\frac{T}{2} \right)}{T \sqrt{\frac{1}{12(n-2)}}}, T = \left(\sum_{j=2}^n j g_j \right).$$

The problem...



Outline - BM

- What is Brownian motion?
- When might characters evolve in a Brownian-like way?
- Simulating Brownian motion on trees
- Independent contrasts

Three Facts Describe Brownian Motion

- Let $W(t)$ be the value of the character at time t . Then:
 - $E[W(t)] = W(0)$
 - Successive steps are independent
 - $W(t) \sim N(W(0), \sigma^2 t)$

Parameters of BM

- Brownian motion models have two parameters:
 - Θ , the starting value; $W(0) = \Theta$
 - σ^2 , the rate parameter

Evolution might approximate BM...

- Genetic drift
- Random punctuated change
- Selection that is weak relative to the time interval considered
- Selection that changes randomly through time

Simulating BM

- Simulating Brownian motion involves drawing values from normal distributions
- Variance of the distribution depends on σ^2 and t
- Values along adjacent branches are added from the root to the tips of the tree

General form

- Tip data follow a multivariate normal distribution with mean vector Θ and variance-covariance matrix where
- $\text{var}(i) = \sigma^2(d_i)$; d_i = distance from root to tip i
- $\text{cov}(i,j) = \sigma^2(c_{i,j})$; $c_{i,j}$ = shared path of tip i and j

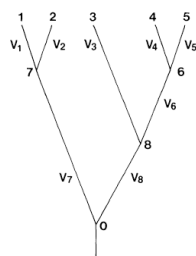


TABLE 1
THE FOUR CONTRASTS EXTRACTED FROM THE PHYLOGENY
SHOWN IN FIGURE 9, EACH WITH ITS VARIANCE, ALL
COMPUTED USING STEPS 1-4 IN THE TEXT

CONTRAST	VARIANCE
$X_1 - X_2$	$v_1 + v_2$
$X_4 - X_5$	$v_4 + v_5$
$X_3 - X_6$	$v_3 + v_6$
$X_7 - X_8$	$v_7 + v_8$

where

$$X_6 = \frac{v_4 X_5 + v_5 X_4}{v_4 + v_5}$$

$$v_6' = v_6 + v_4 v_5 / (v_4 + v_5)$$

$$X_7 = \frac{v_2 X_1 + v_1 X_2}{v_1 + v_2}$$

$$v_7' = v_7 + v_1 v_2 / (v_1 + v_2)$$

$$X_8 = \frac{v_6' X_3 + v_3 X_6}{v_3 + v_6'}$$

$$v_8' = v_7' + v_3 v_6' / (v_3 + v_6')$$

(Felsenstein 1981)

What are contrasts?

- Each standardized contrast is telling us something about the **RATE** of evolution
- The contrasts have a close relationship with σ^2 , the rate parameter from BM
- The sum of the squared contrasts divided by n gives the ML estimate of σ^2

Using and Interpreting Contrasts

- Independent contrasts should be thought of as vectors
- They summarize information about the **amount** and **direction** of evolution at each node in the tree
- Standardized contrasts provide information about the **rate** of evolution

Using and Interpreting Contrasts

$$\hat{\sigma}^2 = \frac{\sum s_i}{n-1}$$

- This provides an unbiased estimate of evolutionary rate
- The expected value of this estimate is equal to the actual rate parameter
- The maximum likelihood estimate of the rate parameter is biased

Character correlations

- Calculate independent contrasts for two characters, x and y
- Carry out a regression analysis of y on x with **no intercept** (force regression line through the origin)
- $P < 0.05$, then reject the null hypothesis of no evolutionary correlation

Outline

- Calculating the likelihood for a single character evolving under a BM model
- Alternative models for continuous character evolution
- Multivariate character evolution

Likelihood for continuous characters on trees

- Given phylogeny, measurements of character y for each tip (y_i)
- Choose a rate parameter σ^2 and mean Θ
- Calculate the phylogenetic variance-covariance matrix for the tree \mathbf{V}

Likelihood for continuous characters on trees

- $y_i \sim \text{MVN}(\Theta, \sigma^2 \mathbf{V})$
- Determine the probability of drawing the vector of y_i from the MVN distribution with mean Θ and vcv $\sigma^2 \mathbf{V}$

Analytic Solution for MLE

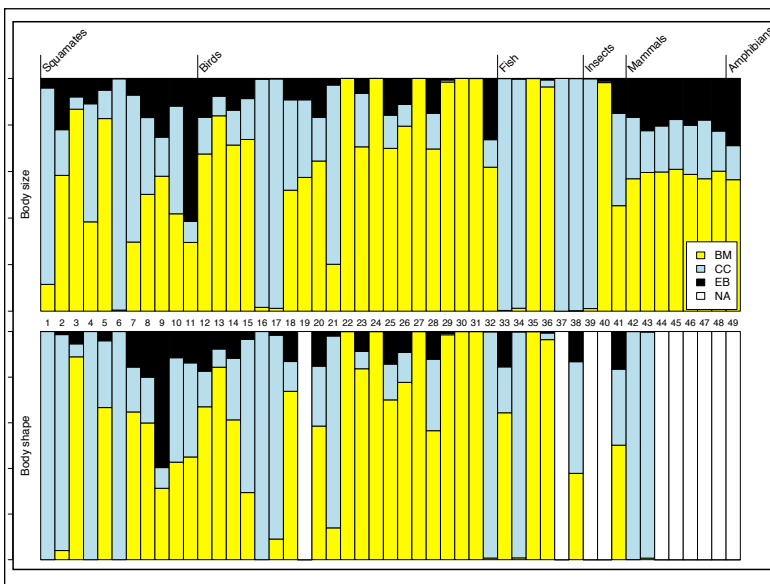
$$\hat{\sigma}^2 = \frac{(\mathbf{x} - \hat{\mathbf{a}}\mathbf{1})' \mathbf{C}^{-1} (\mathbf{x} - \hat{\mathbf{a}}\mathbf{1})}{n}$$

$$\hat{\mathbf{a}} = [(\mathbf{1}' \mathbf{C}^{-1} \mathbf{1})^{-1} (\mathbf{1}' \mathbf{C}^{-1} \mathbf{X})]'$$

\mathbf{x} = vector of trait values, n = number of species,
 \mathbf{C} = coancestry matrix (shared path lengths)

Three Models for Phenotypic Evolution

- Single Rate (SR)
- Early Burst (EB)
- Constant Constraints (CC)



What if you have more than one character?

- All characters for all species are drawn from a multivariate normal distribution
- $V = R \otimes C$
- \otimes is the *Kronecker product*

Outline

- Example of the problem
- The Mk Model
- Alternatives to Mk

The Mk Model

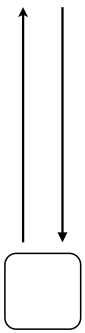
- Simple model for the evolution of discrete variables
- Constant instantaneous probability of changing between any two character states out of k possibilities
- Described by **rate matrix Q**:

$$Q = \alpha \begin{bmatrix} 1-k & 1 & \dots & 1 \\ 1 & 1-k & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 1-k \end{bmatrix}$$

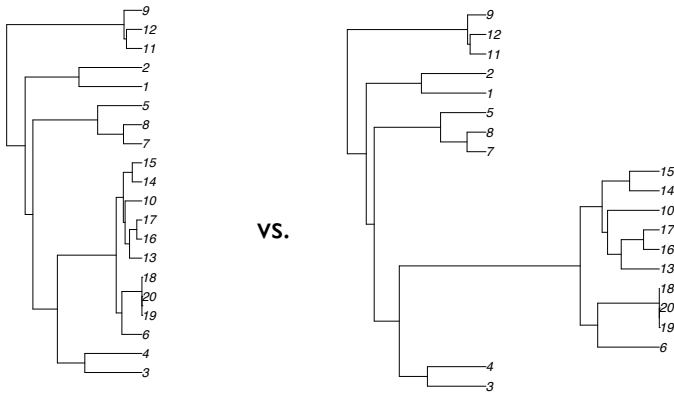
In general:

$$L_{(\tau)} = \sum_{r=1}^{k^{n-1}} \Pr(R_r | \tau)$$

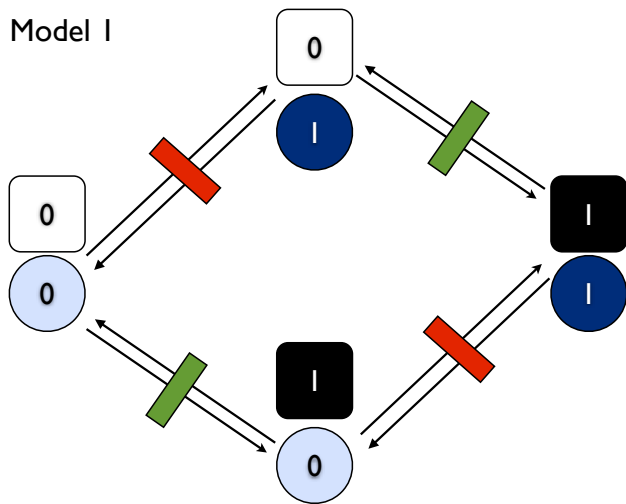
Alternatives to Mk



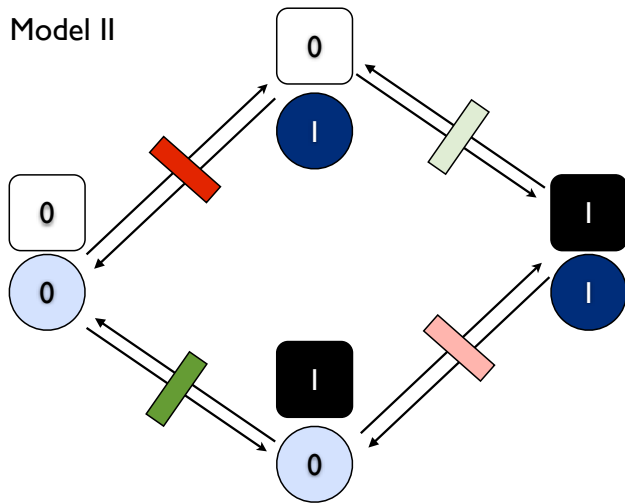
$$Q = \begin{bmatrix} -\alpha & \alpha \\ \beta & -\beta \end{bmatrix}$$



Model I



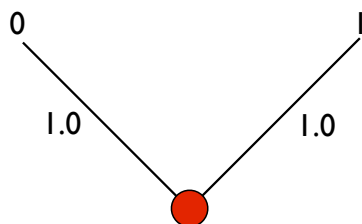
Model II



Reconstructing Ancestral Character States

- Goal: to make inferences about the states of characters in ancestral lineages
- Usually make these inferences at nodes in the tree
- These inferences can be made using maximum likelihood

Discrete Ancestral State Reconstruction



$$L = P(d \mid m) = 0.250$$

$$L(a=0 \mid d, m) = 0.125$$

$$L(a=1 \mid d, m) = 0.125$$

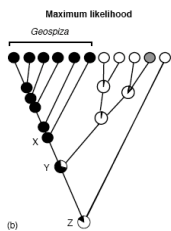
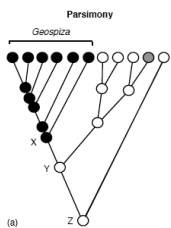
Model: Mk with $q_{01} = q_{10} = 8$

Continuous Characters

- Can also reconstruct ancestral characters using likelihood
- Assume characters are evolving under a Brownian motion model

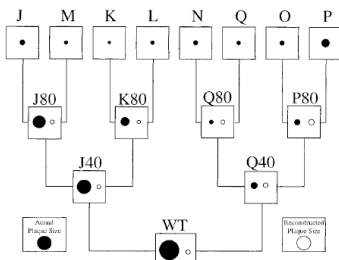
$$\Pr[y = x] = \frac{1}{\sigma\sqrt{2\pi t}} \exp\left(-\frac{(x - \theta)^2}{2\sigma^2 t}\right)$$

Ancestral State Reconstruction



- Lesson 1: always quantify uncertainty

Ancestral State Reconstruction



- Lesson 2: you can't see trends; results are model-dependent

Two New Models

- The
 - Int
 - co
- What is Felsenstein up to now?



General approach

- Model evolution of liabilities l_i
- These evolve under a Brownian motion model with some variance-covariance matrix, but cannot be observed
- Use likelihood or Bayesian MCMC to fit model parameters

Advantages

- Fewer parameters
 - Discrete Markov, 10 variables, >1000 parameters
 - Threshold model, $p(p+1)/2 = 55$ params
- Characters that just changed are more likely to change back.
- Allows for polymorphism

Advantages

- Can estimate both within- and among-species variances and covariances
- Could potentially connect this even better to data collected within species
- Can try to reconstruct selection vectors necessary to get pattern

Disadvantages

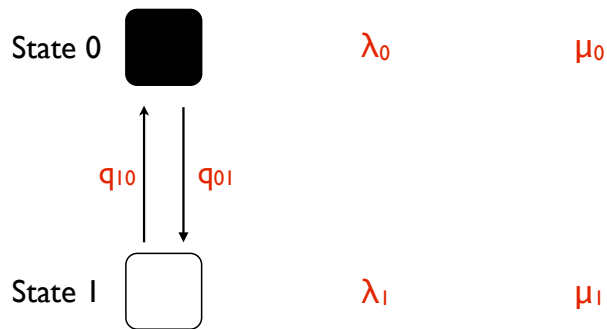
- Not really independent contrasts in the same sense as originals
- Cannot obtain a set of standardized independent contrasts
- Requires iterative method, fancy ML estimation of the two covariance matrices

ML test of multiple rates

- Compare models where some parts of the tree have distinct speciation and/or extinction rates
- If you have likelihoods and the number of parameters, you can compare these models

The BiSSE Model

Speciation rate Extinction rate



Biogeography

- Goals of biogeographic studies
- Classic approaches
- Newer model-fitting methods for biogeography

