Modeling Discrete Character Evolution
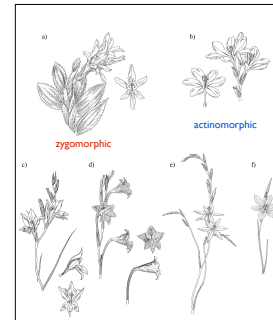
April 6, 2009

Outline

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

Outline

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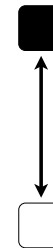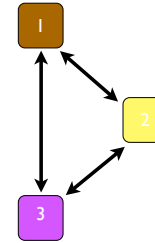- The Mk Model
- Alternatives to Mk


Flower Shape

- Flowers arranged along a single plane of symmetry (zygomorphy)
- Radially, with multiple axes of symmetry (actinomorphy)
- Early angiosperms were actinomorphic
- Hypothesis: Zygomorphy arose multiple times; reversals rare

---

Outline

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

Mk Model

- This is a general version of the GTR model for sequence evolution
- Characters can exhibit one of a set of possible *states*
- Transition rates among these states is governed by a rate matrix (Q)

A (G C T diagram)
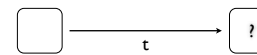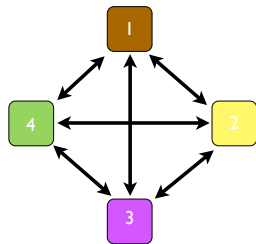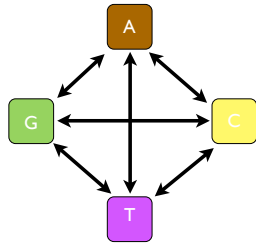
What is the probability distribution
of the character after time t?

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:

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

The Mk Model

$$P_{ii}(t) = \frac{1}{k} + \frac{k-1}{k}e^{-k\alpha t},$$

$$P_{ij}(t) = \frac{1}{k} - \frac{1}{k}e^{-k\alpha t},$$

The Mk Model

- Generalized JC model (Mk with k=4)
- At equilibrium, stationary distribution of characters is [1/k, 1/k, ... , 1/k]
- Described by transition matrix
- m = exp(Qt)

Example
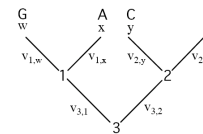
- Consider a three-state character evolving under the Mk model with α = 0.1
- If the character starts out at state 0 at time t = 0, what is the probability that the character will be in state 1 at time t = 10?

$$P_{ii}(t) = \frac{1}{k} + \frac{k-1}{k}e^{-k\alpha t},$$

$$P_{ij}(t) = \frac{1}{k} - \frac{1}{k}e^{-k\alpha t},$$

Matrix Exponential

$$e^X = \sum_{k=0}^{\infty} \frac{1}{k!}X^k$$

$$L^i(\tau) = \text{Prob}\left(\begin{array}{c}\text{tree}\end{array}\right) + \text{Prob}\left(\begin{array}{c}\text{tree}\end{array}\right)$$

$$+ \text{Prob}\left(\begin{array}{c}\text{tree}\end{array}\right) \bullet \bullet \bullet + \text{Prob}\left(\begin{array}{c}\text{tree}\end{array}\right)$$

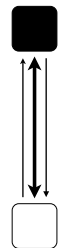$$L^i_{(\tau)} = \sum_r^{4^{n-1}} \Pr(R^i_r \mid \tau)$$

In general:

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

Outline

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

Alternatives to Mk

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

$$Q = \alpha \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix}$$

Alternatives to Mk

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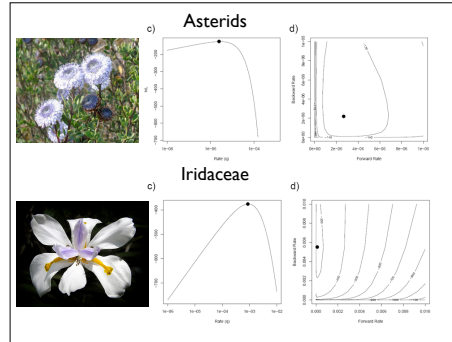
symmetric model

asymmetric model

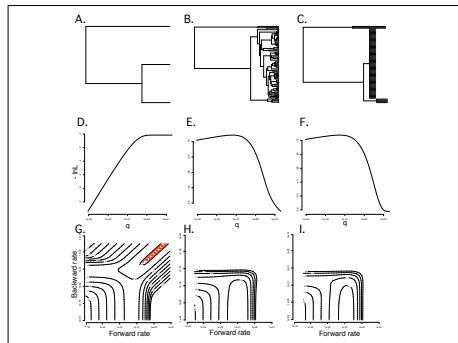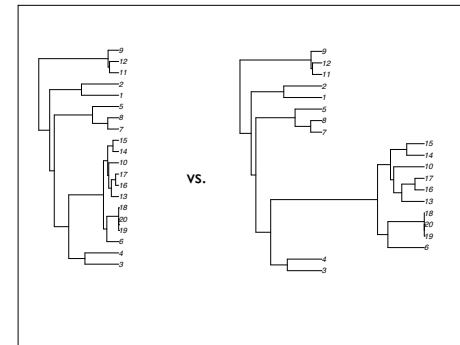$$Q_A = \alpha \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} \qquad Q_B = \begin{bmatrix} -\alpha & \alpha \\ \beta & -\beta \end{bmatrix}$$

Calculate likelihoods under models A and B
Compare using AIC or LR test

Asterids

Iridaceae

Iridaceae

- Model A (Mk model): lnL = -375.7
- Model B (asymmetric model): lnL = -190.8
- Carry out a test. What do you conclude?

vs.

Bayesian Approach

- Newer Bayesian approach from Pagel et al.
- Implemented in the program BayesTraits
- Estimates the posterior distribution of various rate parameters (from the Q matrix)
- May be useful for large datasets

Other models

- Can also model rate variation across lineages and/or through time
- Easiest way to do this is by transforming the branches of a tree

Character Mapping

- Historical approaches to this problem rely on reconstructions
- Map character state changes on branches in the tree, perhaps using parsimony
- Count them in various ways
- Such methods rarely have nice statistical properties, but can be useful in a pinch

Concentrated Changes Test

- Developed by W. Maddison (1990)
- Works for two discrete 0/1 characters
- Reconstruct character 2 on the tree using parsimony
- Reconstruct character 1, and fill in the table
- Decide if the result is unusual given all possible reconstructions of character 1

State of character 2	# of changes in character 1
0	
1	

SIMMAP *Software*

- Software by Bollback
- Can generate character state reconstructions on a tree stochastically
- Simulates the Mk model backwards

Outline

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