Continuous characters

- ML ancestral states for continuous characters under BM are equivalent to ancestral states estimated under weighted squared change parsimony

- Choose the ancestral states that minimize the amount of squared change on the tree

---

FIG. 6. Phylogeny and ecological state (ecomorphs) of Puerto Rican *Anolis* lizards. The unknown state of the second-oldest ancestor is indicated by the question mark. Tree topology and branch lengths are from Losos (1990a). A recent molecular phylogeny of the Caribbean anoles is not in full agreement with the tree shown, but the differences are only weakly supported in the DNA estimate (J. B. Losos, pers. comm. 1996).

Maximum likelihood

- Find model parameters that maximize the likelihood of obtaining the data

- Calculate $P(a=x_i \mid model, tree, data)$

- Can generate confidence intervals based on t-distribution (Schluter 1997)

Ancestral State Reconstruction

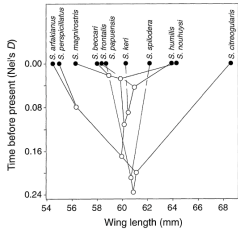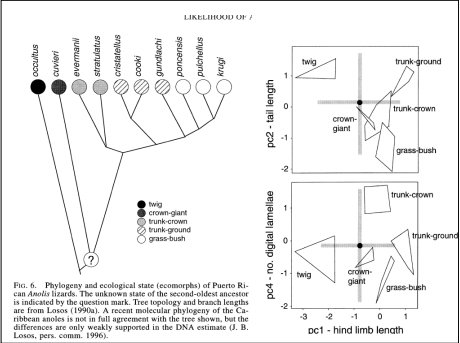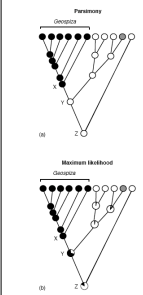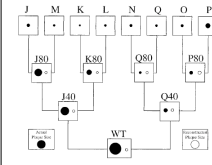- Lesson 1: always quantify uncertainty

---

FIG. 8. Wing lengths of Australian and New Guinea scrubwrens (*Sericornis*) and of their ancestors. The tree is the same as in Figure 3. Shaded lines span the support limits for each ancestor state. Marginal distributions were computed separately for each ancestor one at a time.

Ancestral State Reconstruction

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

Summary

- We can test for character correlations, or other patterns of interest, with discrete characters
- It is possible to estimate ancestral character states and incorporate uncertainty

Roots of the Problem

- Gould pondered the connection between micro- and macroevolution
- Identified the problem of stasis: why do we see so little change in so many groups, even over long time periods?
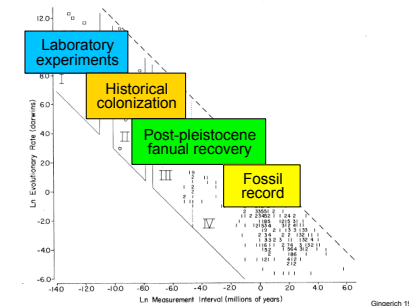
PROFESSOR STEPHEN JAY GOULD

## Comparative Methods and Quantitative Genetics

April 14, 2009

A General Pattern in Macroevolution

Laboratory experiments

Historical colonization

Post-pleistocene fanual recovery

Fossil record

Gingerich 1983

Quantitative Genetics and Macroevolution

- Comparative biology has the opportunity to test long-term prediction of quantitative genetic models
- How much change would we expect if we extrapolate from present-day trends?
- How much change do we actually see in the fossil record?

Lynch (1990) Am. Nat.

Rate of evolution is slow even compared to the lower limit of the neutral expectation
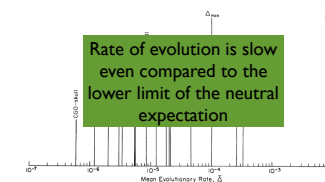
What is the point?

- We can learn things by comparing results from comparative biology to microevolutionary models

- This helps bridge the gap between micro- and macroevolution

---

Two New Models

- Th...
- Int...
  co...

What is Felsenstein up to now?

Resolving the Paradox of Stasis: Models with Stabilizing
Selection Explain Evolutionary Divergence
on All Timescales
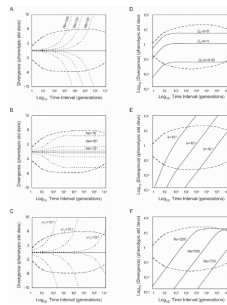
Suzanne Estes[*] and Stevan J. Arnold[†]

Threshold models

- Problem with analyzing quantitative characters:
  - Number of parameters increase too fast with multivariate data sets
  - Biologically unrealistic patterns; easy to reverse states even after long periods of time

---

Figure 6: Example fits of neutral (A), white noise (B), Brownian motion of the optimum (C), displaced optimum (D), moving optimum (E), and peak shift (F) models to the divergence data. Heavy dashed line is each panel represent the 90% confidence ellipse for the divergence data of Gingerich (2001). If model predictions fell within this ellipse, we considered the fit to be data acceptable. To illustrate how each model responds

---

Sewell Wright

- Introduced threshold model for guinea pig toes in 1934
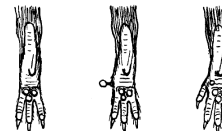

FIGURE 1.—Sketches of left hind feet of guinea pigs. The normal 3-toed condition is shown at the left. At the right is a foot with well-developed little toe and a corresponding plantar tubercle. Between is a grade of imperfect development of the little toe.

Wright 1934 Genetics

Slide 1 (figure):

p characters, n species:

$$L = \text{Prob}(y, \mu, C)$$

$$= \int_{x_{11} \in R_{11}} \int_{x_{12} \in R_{12}} \cdots \int_{x_{1p} \in R_{1p}} \int_{x_{21} \in R_{21}} \cdots \int_{x_{np} \in R_{np}} \phi(x; \mu, C), \quad (2.2)$$

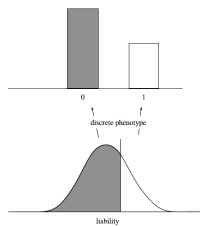Hard to calculate:
Felsenstein suggests MCMC

Figure 1. The threshold model of quantitative genetics, showing the continuous distribution of the underlying liability characters, and the resulting discrete distribution of the observed phenotype.
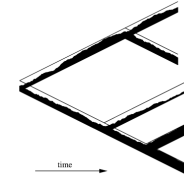
Slide 2 (figure):

Figure 2. A simulation of a threshold model with one character, evolving by Brownian motion of the liability on a simple five-species tree. Each lineage is shaded to indicate the proportion of individuals in state 1 (shown by black shading). The upper two species end up with polymorphism, the middle species nearly all with state 0, and the lower ones nearly all with state 1.

Slide 3:

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

Slide 4:

Testing for correlations with the threshold model

- If liabilities are correlated then the two characters are evolving in a correlated manner
- Can test this using model-fitting approach

Slide 5:

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

Slide 6:

Two New Models

- Threshold models
- Intra- and interspecific contrasts

Contrasts Revisited

- There is no clean line separating phylogenetic trees from intraspecific geneologies
- Comparative methods almost all take species values as fixed and estimated perfectly
- But traits vary in the natural world, and we estimate them with error

---

Calculating Contrasts*

1. Find two tips on the phylogeny that are adjacent (say nodes i and j) and have a common ancestor, say node k
2. Compute the contrast X_i-X_j. This has expectation zero and variance proportional to v_i+v_j

$$C_{ij} = X_i - X_j$$
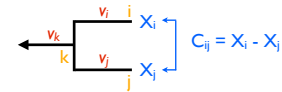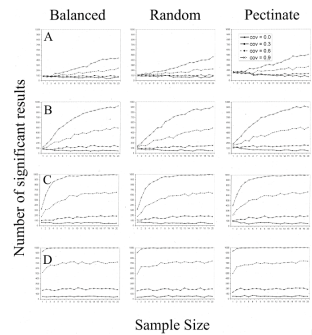

Contrasts Revisited

- Why is there a problem?
- We assume that species means are estimated without error in most comparative methods
- This can lead to statistical problems

Also inflates rates

$$v_{ij} = v_i + v_j \qquad \text{Contrasts}$$

$$v_{ij} = v_i + v_j + \frac{\sigma_i^2}{n_i} + \frac{\sigma_j^2}{n_j} \qquad \text{Actual}$$

This leads to inflated values for independent contrasts, especially towards the tips of the tree
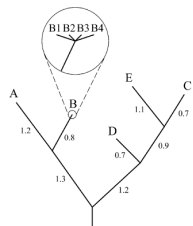

Balanced   Random   Pectinate

Number of significant results

Sample Size

Harmon and Losos 2005

Felsenstein's model

- Character evolution on the tree is described by an evolutionary variance-covariance matrix R
- Remember:

Slide 1

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

Slide 2

Among-species contrasts

$$x = \frac{[1/(v_A + \delta v_L)]x_1 + [1/(v_B + \delta v_R)]x_2}{[1/(v_A + \delta v_L)] + [1/(v_B + \delta v_R)]}.$$

$$\delta v = \frac{1}{[1/(v_L + \delta v_L)] + [1/(v_B + \delta v_R)]}.$$

Similar, but not identical too,
original contrasts

Slide 3

Felsenstein's model

- Characters also vary within species according to a (distinct) variance-covariance matrix W
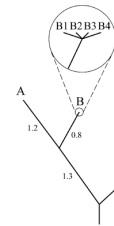- Assumed to be the same for each species in the tree

Slide 4

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

Slide 5

Within-species contrasts

$$c_1 = f_1(x_{B1} - x_{B2})_2$$

$$c_2 = f_2\left[x_{B3} - \left(\frac{1}{2}x_{B1} + \frac{1}{2}x_{B2}\right)\right]_2$$

$$c_3 = f_3\left[x_{B4} - \left(\frac{1}{3}x_{B1} + \frac{1}{3}x_{B2} + \frac{1}{3}x_{B3}\right)\right]_2$$

$$f_m = \sqrt{\frac{m}{m+1}}.$$

Slide 6

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

Two New Models

- Threshold models
- Intra- and interspecific contrasts