Robust Information from Phylogenetic Trees?
Carl Boettiger, UC Davis

Process dynamics look like:

Imagine the evolution of a continuous character trait which may occasionally jump from one optima to another—a large ground-dwelling species becoming smaller tree-dwellers. If we could observe this entire process over many species, it might look like this.

Can we reconstruct it with this…

We see only a snapshot in time. Each species as a replicate from this stochastic process. We detect clusters, but can’t calculate rates.

and a Phylogenetic Tree?

Species trait values don’t represent independent sample paths. The phylogenetic tree introduces correlations. We will rely on this for temporal information. Colors indicate models with different optima (paintings).

Data given the painting

The likelihood of a set of species mean traits is jointly Gaussian even if the model has multiple regimes.

Painting given transition matrix

Exponentiating the transition matrix \( Q \) determines the probability of over all possible paths from state \( i \) to state \( j \). Taking the product over all branches \( i, j \) of the tree gives the probability of the painting given \( Q \).

\[
\mathcal{L}(X) = \prod_{ij} \left\{ e^{Q_{ij}} \right\} P_{ij}
\]

Summary

- Can infer dynamic evolutionary parameters without fossil history
- Handle uncertainty more robustly than current methods

Partioning a Hard Problem

\[
P(\text{painting}) = P(\text{painting} | \text{state})P(\text{state})
\]

Generating paintings from a Markov model rather than as a hypothesis is preferable biologically and statistically. A clever partition of likelihood makes this numerically feasible.

Acknowledgements

- C. Martin, S. Price, P. Wainwright
- P. Ralph, G. Coop

Likelihood Ratio

Comparing likelihoods of two hypotheses: if \( X \) is the observed data and \( \theta \) the parameter then

\[
P(\text{data} | \theta_1, \theta_2)P(\theta_2)P(\theta_1)
\]

The natural log transform is suitable for comparing likelihoods for non-nested models because the difference in natural log likelihoods has a \( \chi^2 \) distribution.

\[
\chi^2 = 2 \left( \ln L_1 - \ln L_2 \right)
\]