Modelling rate distributions using character compatibility: implications for morphological evolution among fossil invertebrates

Peter J. Wagner*
National Museum of Natural History, Smithsonian Institution, Washington DC 20013, USA
*wagnerp@si.edu

Rate distributions are important considerations when testing hypotheses about morphological evolution or phylogeny. They also have implications about general processes underlying character evolution. Molecular systematists often assume that rates are Poisson processes with gamma distributions. However, morphological change is the product of multiple probabilistic processes and should theoretically be affected by hierarchical integration of characters. Both factors predict lognormal rate distributions. Here, a simple inverse modelling approach assesses the best single-rate, gamma and lognormal models given observed character compatibility for 115 invertebrate groups. Tests reject the single-rate model for nearly all cases. Moreover, the lognormal outperforms the gamma for character change rates and (especially) state derivation rates. The latter in particular is consistent with integration affecting morphological character evolution.

Keywords: character evolution; compatibility; lognormal distribution; gamma distribution; information theory

1. INTRODUCTION

Palaeontologists are interested in rates of character state change both for testing macroevolutionary hypotheses and for inferring phylogeny. However, palaeontologists have paid much more attention to rate variation over time and among clades than to rate variation among characters. Indeed, most rate and systematic studies tacitly assume a single rate for all characters. We have not yet explored whether there are any general rules for rate distributions, or even whether single-rate models are poor ones.

Molecular systematists often model rate variation with gamma distributions. This assumes a collection of Poisson processes with different ‘waiting times’ between events [1]. However, palaeontologists use only fossilizable morphology. Morphological change is probably the product of waiting time and probabilistically varying selective forces. Moreover, biomechanics, development and genetics probably create hierarchical interactions among morphological characters [2]. Rates for any one character thus should partially reflect both the number of associated characters and the probabilities of those characters accommodating a change. If overall rates are products of independent probabilistic processes [3] and/or hierarchical linkages [4], then morphological rates should fit lognormal distributions.

Palaeontologists have combined simulations with the morphological disparity of character matrices to assess general rates of change (e.g. [5]). Systematists have used the related concept of compatibility to contrast general rates for individual characters (e.g. [6]) (see the electronic supplementary material for a contrast of disparity and compatibility). Here, I unite both traditions to assess how well three basic models of character evolution (single rates, gamma distributions and lognormal distributions) predict fossil morphologies.

2. MATERIAL AND METHODS

(a) Compatibility, steps and rates

Yang [1] suggests using steps implied by parsimony trees to establish rate distributions. However, when simulated matrix structure matches that of real data, parsimony steps blur rate classes [7]. Instead, I use character compatibility to approximate steps. A compatible character pair has no necessary homoplasy: a binary pair combining for 00, 01 and 11 might not have homoplasy but a pair also including combination 10 must have homoplasy [6,8]. Simulations corroborate the suggestion that compatibility decreases as frequencies of change (and homoplasy) increase, both for whole matrices [7] and for individual characters [9].

State distributions also affect the probability of compatibility. Suppose two binary characters change only once. For character A, only one taxon possesses state 1. State 1 cannot be paired with two states in any other character, and thus cannot create incompatibility. Conversely, character B changes in the middle of the phylogeny so that half of the taxa have state 0 and other half have state 1. Both states 0 and 1 for character B now have four opportunities to be paired with a second state for other binary characters. In simulations, compatibility given X steps decreases as states become more ‘evenly’ distributed among taxa (see electronic supplementary material).

This study uses simulations to inversely model the probability of observed compatibilities and state distributions given the rates of character state change. For a clade of S taxa with N[0] characters, each simulation:

- evolves a tree of S taxa;
- evolves N[0] characters (retaining the observed state numbers and missing entries as the data) until the observed compatibility is reached;
- iteratively removes simulated characters matching the compatibility of observed characters;
- re-simulates the ‘matching’ character with 1, . . . , S steps; and
- tallies how often we find observed compatibility and state distributions given 1, . . . , S steps.

There are three critical differences between this approach and permutation tests assessing character compatibilities (e.g. [10–12]). First, phylogeny underlies the character state distributions even under high rates of change (part 1). Second, each character is compared with a ‘remaining’ matrix having the same compatibility as the observed ‘remaining’ matrix (parts 2 and 3). Finally, results are tied to specific amounts of change (parts 4 and 5).

Now, we have the likelihoods of 1, . . . , S changes (steps) given the observed compatibility for each character. However, model evaluation requires numbers of characters with X steps (see below). This, in turn, requires posterior probabilities rather than likelihoods. Assuming flat priors for step numbers, the probability of character C having X steps is:

\[
p(\text{compatibility of C|X steps}) = \sum_{Y=1}^{\infty} p(\text{compatibility of C|Y steps})
\]
The estimated number of characters with \( X \) steps now is:

\[
\sum_{i=1}^{\text{characters}} p(\text{compatibility of C}_i | \text{X steps})
\]

Character evolution models, such as the Mk model [13], describe rates of character change and rates of state derivation. This distinction is critical when characters have different numbers of states. Consider simple transition matrices for two-state character A and a three-state character B, where the off-diagonals give \( p(\text{state derivation}) \) and the diagonal gives \( p(\text{stasis}) \) (i.e. \( 1 - p(\text{change}) \)):

\[
\begin{array}{c|ccc|c}
\text{states} & 0 & 1 - \delta_A & 1 & \text{versus} \\
0 & 1 - \delta_B & \frac{\delta_B}{2} & \frac{\delta_B}{2} & \frac{2}{2} \\
1 & \delta_B & 1 - \delta_B & \frac{\delta_B}{2} & \frac{2}{2} \\
2 & \frac{\delta_B}{2} & \frac{\delta_B}{2} & 1 - \delta_B & \frac{2}{2}
\end{array}
\]

If \( \delta_A = \delta_B \), then the rate of state derivation is necessarily lower for B than for A; conversely, if the rates of state derivation are the same, then \( \delta_B = 2\delta_A \). An upshot of this is that rates of character change and rates of state derivation need not fit the same models.

These analyses assume equiprobable state transitions. The issue of biased-state derivation requires a phylogenetic context, if only to assess the primitive state.) Thus, models are not fitted using how frequently particular states are derived, but how frequently we derive the first or second ‘other’ state given any ancestral state. There is at least one derivation for each ‘other’ state (e.g. the two ‘other’ states for character B above must be derived once each). At four steps, we have two additional derivations, so for each ‘other’ state:

\[
p(\text{0 additional derivations}) = p(\text{2 additional derivations}) = 0.25; \quad p(\text{1 additional derivation}) = 0.50.
\]

As there are two ‘other’ states, we expect \( 2 \times 0.25 = 0.50 \) states to be derived either once or three times in total, and \( 2 \times 0.50 = 1.0 \) to be derived twice the total. The number of states from each character expected to be derived \( Y \) times now is the sum of \( p(\text{X derivations} | \text{X steps}) \times p(\text{X steps} | \text{data}) \) for \( X = 1, \ldots, S \). This is summed over all characters. Analyses then use that final summation to evaluate different models (see electronic supplementary material).

(b) Rate distribution models

The single-rate model varies one parameter: mean rate, \( \delta \). Gamma distributions use a scale and a shape parameter to vary rates around the mean rate \( \delta \). Following Yang [1], I set the scale equal to shape, which leaves one freely varying parameter. Increasing the size/shape parameter decreases the rate variation. The lognormal uses an order of magnitude increase/decrease given the deviation from the mean rate. Thus, \( p(\text{rate} = 2\delta) = p(\text{rate} = \delta/2) \). Thus, both the gamma and lognormals vary two parameters: mean rate \( \delta \) and the distribution parameter. Analyses then divide the distributions into four equal areas [1], with the midpoint of each quadrant used for the rate relative to \( \delta \).

Characters are not assigned to particular rate classes. Instead, likelihoods contrast the reconstructed and expected number of characters with \( 1, \ldots, S \) changes given \( N'/4 \) characters with rates 1, 2, 3...
and 4, where $N$ is a hypothetical ‘true’ number of characters including the varying characters ($N_v$) plus some invariant ones. Note that $N_v$ is an outcome of the best hypotheses rather than a freely varying parameter. The procedure is repeated for state derivations (see electronic supplementary material, appendix A).

Distribution likelihoods are

$$L_{\text{rates|observations}} \propto \sum_{i=1}^{S} f_{\text{observations|}i \text{ steps}} \times P_{i \text{ steps|rates}}.$$ 

I then use Akaike’s modified information criterion (AICc [14]) to rescale model log-likelihoods given equal parameter and sample size ‘efforts’ for all three models:

$$\text{AICc} = -2 \ln L_{\text{ML hypothesis for model}} + 2K \left( \frac{N}{N - K - 1} \right),$$

where $K$ is the number of parameters (1 for single-rate, 2 for lognormal and gamma) and $N$ is the number of data points (i.e. observed characters or derived states). Each model is evaluated by the differences between its AICc and the overall best AICc. For the ‘null’ single-rate model, I use Akaike’s weights [15]:

$$w_i = \frac{e^{-\Delta \text{AICc}_i}}{\sum_{\text{model}=1}^{2} e^{-\Delta \text{AICc}_{\text{model}}}}.$$ 

This rescales data probabilities given differences in model parameters and then relative to the sum of rescaled data probabilities for all three models. I use AICc differences ($\Delta \text{AICc}$) to contrast the gamma and lognormal hypotheses.

(c) Data

I analyse 115 published character matrices of fossil echinoderms, molluscs and trilobites (see electronic supplementary material, appendix B). For most matrices, I exclude outgroup taxa (especially hypothetical ancestors). Analyses exclude dependent characters, i.e. those describing particular conditions of features absent on some clade members. These require an extra layer of analysis where change is conditioned on the state of a particular independent character. Dependent characters also require allowing for zero steps as two states might reflect two derivations of the independent character. Finally, polymorphic characters are set to states that maximize the compatibility of the character.

3. RESULTS

Single-rate models explain character compatibility patterns poorly relative to distributed rate models (figure 1). Half of the character rate examples have $w_{\text{Rate}} < 10^{-10}$ and half of the state derivation examples have $w_{\text{Rate}} < 10^{-9}$.

For rates of character change, the lognormal model outperforms the gamma model in 78 of 115 cases (figure 2). The pattern is much stronger for...
echinoderms (23 of 29 cases) and molluscs (29 of 39 cases) than for trilobites (26 of 47 cases). For rates of state derivation, the lognormal performs best across all three groups (108 of 115 cases).

4. DISCUSSION

We can clearly reject the idea of single-rate models of morphological character change for the majority of invertebrate groups examined here. The few cases where we cannot do so use few taxa and/or characters and thus implicate sample size. Allowing for distributed rates might alter interpretations of macroevolutionary studies showing altered mean rates (δ) over phylogeny. For example, δ is lower for derived rostroconchs than for stem rostroconchs [16]. However, these analyses indicate that the shift reflects decreases to the slower rate classes while the highest rates remain little changed (see electronic supplementary material, appendix A).

Basic phylogenetic inference methods (e.g. [17,18]) assume single-rate models of character change. Simulations demonstrate that rate heterogeneity confounds such methods (e.g. [19,20]). Likelihood and Bayesian methods (e.g. MrBayes; [21]) allow distributed rates but currently consider only gamma distributions (e.g. [22]). The extent to which distributed rates in general and lognormal distributions in particular alter phylogenetic inferences of fossil taxa requires examination.

The preponderance of lognormal rate distributions suggests that morphological evolution is more than a collection of Poisson processes. The two explanations, i.e. multiple independent processes and hierarchical linkage/integration, are not mutually exclusive. However, ‘independent processes’ might best pertain to when characters change, whereas ‘integration’ might best pertain to how characters change (state derivation). Of course, rates of state derivation will affect rates of change, and vice versa. However, for several clades, gamma better describes rates of change whereas lognormal better describes rates of derivation.

This is particularly true for trilobites (figure 2). This suggests that simple stochastic processes affect when change happens, but some other set of rules (such as integration) affect what change can occur. Proper testing of these ideas is beyond the scope of this paper. Nevertheless, the dominance of lognormal rate distributions over gamma rate distributions shows that palaeontological data leave more evidence of underlying processes than previously suggested.

I thank D. H. Erwin, J. D. Marcot, K. E. Sears and S. K. Lyons for discussion.


