

Morphology Review

Wright Lab Y'all

Today

Upwards of 99% of species on Earth have gone extinct. The oldest DNA samples that exist are 560-780,000 years old (cite horse paper), and prospects are dim for gathering DNA older than 1-200,000 years old. Therefore, in order to observe ancient biodiversity, and understand the evolutionary history of life on Earth, we must make use of the fossil record. Many methods for understanding paleoecology, macroevolution, and trait evolution rely on phylogenetic trees.

For estimating phylogenetic trees from morphological data, two paradigms have historically dominated. Maximum parsimony is an optimality criterion by which a tree with the fewest changes is preferred. Bayesian methods, in particular the Mk model of Lewis (2001), estimate phylogenetic trees under a dynamic model of evolution. The two methods make strikingly different assumptions about the underlying evolutionary process that lead to the data.

Like parsimony, probabilistic models make a number of assumptions about the way morphological data evolved. Paul Lewis introduced the Markovian K -States, or Mk, model of morphological evolution in 2001. The model is so-named because of the assumption that each character is always in one of k known states. One key assumption of Lewis' model is that there is an equal rate of transition between character states. For example, in a binary matrix, there is an equal probability of transitioning from a '0' state to a '1' state. Because this is a continuous-time model, these changes could occur anywhere along a branch, even including multiple changes occurring on a branch.

Unobserved or 'hidden' states violate the assumption that the data are always in k states, which are known to us and observed in our dataset. The Mk model relies on a Q-matrix. If there are hidden states, then the matrix will be the wrong size, inflating transition rates. Beaulieu et al indicate that this may inflate the lengths of branches when hidden states are not accounted for in state-dependent speciation and extinction models. This is also a risk for phylogenetic estimation itself.

The assumption of symmetry has been observed to be violated in about half of datasets observed by Wright, Lloyd and Hillis. Simulation results confirmed that model misspecification in failing to account for this asymmetry severely negatively impacts phylogenetic accuracy. To more adequately model phylogenetic data, the symmetric Beta prior was introduced in MrBayes, and extended in RevBayes (Wright, Johnson, Jenkins, Pett, Heath). Our model places a beta prior on character frequencies to allow variation in the frequencies character

states. For example, if state 0 has a high transition rate to state 1, but state 0 is infrequently observed, that transition will not be common. This model has also been extended to model character state frequency variation in multistate characters using a Dirichlet distribution.

Parsimony notably has difficulty with long branch lengths. In particular, parsimony is statistically inconsistent when we know the dataset has homoplasy. Parsimony prefers not to accept multiple evolutions of a character as an explanation for the data; instead interpreting that similarity as evidence of common descent. The Mk model, by contrast, can accept the evolution of the same character state evolved multiple times, and, in practice, increase the rate of evolution on the tree.