Phylogenetic Inference using `RevBayes`

Discrete Morphology

April M. Wright and Michael J. Landis

Introduction

While molecular data have become the default for building phylogenetic trees for many types of evolutionary analysis, morphological data remains important, particularly for analyses involving fossils. The use of morphological data raises special considerations for model-based methods for phylogenetic inference. Morphological data are typically collected to maximize the number of parsimony-informative characters - that is, the characters that favor one topology over another. Morphological characters also do not carry common meanings from one character in a matrix to the next; character codings are made arbitrarily. These two factors require extensions to our existing phylogenetic models. Accounting for the complexity of morphological characters remains challenging. This tutorial will provide a discussion of modeling morphological characters, and will demonstrate how to perform Bayesian phylogenetic analysis with morphology using `RevBayes`.

Contents

The Discrete Morphology guide contains several tutorials

- Section 1: Overview of the Discrete Morphological models
- Section 2: A simple discrete morphology analysis
- Section 3: Two complex discrete morphology models

Recommended tutorials

The Discrete Morphology tutorials assume the reader is familiar with the content covered in the following `RevBayes` tutorials

- `Rev` **Basics**
- **Molecular Models of Character Evolution**
- **Running and Diagnosing an MCMC Analysis**
- **Divergence Time Estimation and Node Calibrations**

1 Overview of Discrete Morphology Models

Molecular data forms the basis of most phylogenetic analyses today. However, morphological characters remain relevant: Fossils often provide our only direct observation of extinct biodiversity; DNA degradation can make it difficult or impossible to obtain sufficient molecular data from fragile museum specimens. Using morphological data can help researchers include specimens in their phylogeny that might be left out of a molecular tree.

To understand how morphological characters are modeled, it is important to understand how characters are collected. Unlike in molecular data, for which homology is algorithmically determined, homology in a character is typically assessed an expert. Biologists will typically decide what characters are homologous by looking across specimens at the same structure in multiple taxa; they may also look at the developmental origin of structures in making this assessment. Once homology is determined, characters are broken down into states, or different forms a single character can take. The state '0' commonly refers to absence, meaning that character is not present. In some codings, absence will mean that character has not evolved in that group. In others, absence means that that character has not evolved in that group, and/or that that character has been lost in that group. This type of coding is arbitrary, but both **non-random** and **meaningful**, and poses challenges for how we model the data.

Historically, most phylogenetic analyses using morphological characters have been performed using the maximum parsimony optimality criterion. Maximum parsimony analysis involves proposing trees from the morphological data. Each tree is evaluated according to how many changes it implied in the data, and the tree that requires the fewest changes is preferred. In this way of estimating a tree, a character that does not change, or changes only in one taxon, cannot be used to discriminate between trees (i.e., it does not favor a topology). Therefore, workers with parsimony typically do not collect characters that are parsimony uninformative.

In 2001, Paul Lewis introduced a generalization of the Jukes-Cantor model of sequence evolution for use with morphological data. This model, called the Mk (Markov model, assuming each character is in one of k states) model provided a mathemetical formulation that could be used to estimate trees from morphological data in both likelihood and Bayesian frameworks. While this model is a useful step forward, as a generalization of the Jukes-Cantor, it still makes fairly simplistic assumptions. This tutorial will guide you through estimating a phylogeny with the Mk model, and two useful extensions to the model.

1.1 Some math

...

1.2 Some RevBayes

```
n_areas <- 3
```

1.3 Things to consider

...

2 Overview of Discrete Morphology Models

3 Overview of Discrete Morphology Models

? Alright, who did it?

References

Version dated: February 19, 2017