Phylogenetic Comparative Methods
Summer Internship Presentation

James Hayes

Thursday June 30th 2022

Overview

- **Phylogenetics** - The study of the evolutionary history and evolutionary relationships within or between groups of organisms.
- **Phylogenetic tree** - Diagram that shows evolutionary relationships among species.
- Often the actual evolutionary relationship between species is not known, so the phylogenetic tree must be inferred.
- We can use phylogenetic trees to implement phylogenetic comparative methods allowing us to investigate evolutionary hypotheses.
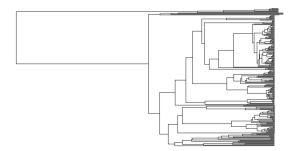
Introduction

Figure: The Mammalian tree of life, compiled using data from: Vincze, O., Colchero, F., Lemaître, JF. et al. Cancer risk across mammals. Nature 601, 263–267 (2022). https://doi.org/10.1038/s41586-021-04224-5

A Motivating Example - Peto's Paradox

- At the most basic level, cancer is caused by certain cell mutations.
- Therefore, we would intuitively expect that the higher an organism's body mass, the more cells it has and the higher the probability that it develops cancer.
- Similarly, we would expect that the longer an organism's lifespan, the more time for a mutation to occur and the higher the probability that it develops cancer.
- This has been observed within humans.
- But at an interspecies level, evidence hasn't been found to support this seemingly intuitive hypothesis - Peto's paradox.
- We could thus employ phylogenetic comparative methods to examine, for example, the evolutionary relationship between the traits of body mass and cancer mortality.

Modelling Evolution of Continuous Variables

- To implement comparative methods we need to assume an evolutionary model that represents the process by which the trait(s) evolved.
- We will briefly discuss two of the main models of evolution used for continuous variables:
  - **Brownian Motion**
    - This is the model of evolution that is assumed for the two comparative methods discussed in this presentation.
  - **The Ornstein-Uhlenbeck Process**

Brownian Motion

- The Wiener process is a stochastic process that approximates the Brownian motion of a particle.
- In this process, the particle takes a large number of small steps on an axis, each of which displaces the particle by a random amount.
- The steps are independent of one another and identically distributed with mean displacement 0 and variance of displacement σ^2.
- Then by the Central Limit Theorem, after sufficiently many steps, the sum of the displacements will be approximately normally distributed with mean 0 and variance $n\sigma^2$, where n is the number of steps.

Brownian Motion

- We can approximate an evolutionary process by saying it follows a Brownian motion with variance σ^2 per unit time.
- Then the net change along a branch of length t is drawn from a normal distribution with mean 0 and variance $\sigma^2 t$.
- Because the small steps that make up the displacement along a branch are independent, the displacements in different branches are independent.

Ornstein-Uhlenbeck Process

- The Brownian motion model, however, ignores the possibility of natural selection occurring on the trait.
- In the case where a number of species are all being selected towards a common optimum, the Ornstein-Uhlenbeck process can be used to model the evolution of the trait.
- This process is Brownian motion but with a force continually acting on the particle towards a point.

Phylogenetic Comparative Methods

- We'll discuss the example of testing whether there is a relationship between two traits.
- **Problem**: Due to shared lineages, species don't provide independent data points for analysis.
- We will need to account for this.

Comparative Methods: Phylogenetically Independent Contrasts (PICS)

- The logic behind PICs is to transform the original tip data into values that are statistically independent and identically distributed.
- We can do this by taking a series of contrasts between character values at the tips.
- Under a Brownian motion model, a contrast is due to displacement that has occurred since the two species split - ie the non-shared part of their evolutionary histories.
- Hence, the contrasts are statistically independent.
- We can then standardize the contrasts by dividing each by the square root of its variance.
- We then perform ordinary least squares regression on the contrasts.
- PICs is a special case of the next method we will discuss, assuming Brownian motion and with Pagel's λ equal to 1.

Phylogenetic Generalized Least Squares (PGLS)

- Perform a weighted regression - PGLS is a special case of generalised least squares.
- **Key insight: Under a Brownian motion model of evolution, expected covariance of the residuals will be related to the shared evolutionary history between species.**
- So the expected covariance between two species is proportional to the shared branch length of those species on the tree.
- So in practice, PGLS penalizes points for their shared evolutionary history with other species.

Phylogenetic Generalized Least Squares (PGLS)

- The method just described assumes the phylogeny accurately describes the evolution of the traits.
- But what if the phylogeny or the assumed evolutionary model is inaccurate?
- We would like if out analysis could incorporate some sort of measure of the phylogenetic signal - the extent to which closely related species truly resemble one another.
- One such measure is Pagel's λ, which we can incorporate into the model.
- λ is a mulitplier of the off-diagonal elements of the covariance matrix.
- λ can be thought of as a scaling factor on the internal branches of the tree.
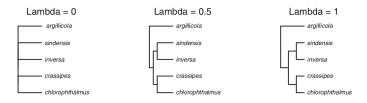
Phylogenetic Generalized Least Squares (PGLS)

Figure: Symonds, M.R.E., Blomberg, S.P. (2014). A Primer on Phylogenetic Generalised Least Squares. In: Garamszegi, L. (eds) Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-43550-2_5