

Comparative Methods

Brian O'Meara

2017-02-17

Contents

| | | |
|-----------|---|-----------|
| 1 | Introduction | 5 |
| 2 | First steps | 7 |
| 3 | Getting data and trees into R | 9 |
| 4 | Visualizing data before use | 11 |
| 5 | Dull model testing | 13 |
| 6 | Continuous traits | 15 |
| 7 | Brownian Motion and Correlations | 17 |
| 8 | Discrete Traits | 19 |
| 8.1 | Objectives | 19 |
| 9 | Diversification | 21 |
| 10 | SSE methods | 23 |
| 11 | RAxML | 25 |
| 12 | Gene Tree Species Tree | 27 |
| 13 | Dating | 29 |
| 14 | Build a method | 31 |
| 15 | Placeholder | 33 |
| 16 | Appendix | 35 |

Chapter 1

Introduction

Chapter 2

First steps

Chapter 3

Getting data and trees into R

Chapter 4

Visualizing data before use

Chapter 5

Dull model testing

Chapter 6

Continuous traits

Chapter 7

Brownian Motion and Correlations

Chapter 8

Discrete Traits

8.1 Objectives

By the end of this chapter, you will:

- Understand how to incorporate rate heterogeneity in discrete trait models
- Be able to explain how to test hypotheses about univariate trait evolution.

Many traits can be thought of as discrete traits: a DNA site comes in ATGC, proteins have one of 20 amino acids, some animals have functional eyes and others do not, some plants are woody and others are herbaceous. This is nearly always an approximation. Think of something like limbs: they seem distinct enough that we even name some groups by their count: tetrapods, hexapods. Except that when we look closely enough, it becomes fuzzy: insect mouthparts are derived from limbs, for example, so should we count these highly modified limbs as limbs (and if not, where in evolution have they become sufficiently modified to no longer count? And are nymphalid butterflies tetrapods under that definition yet?). Are modern whales thought to have four limbs, even though two are extremely vestigial? Often for neontologists problematic organisms with intermediate counts are conveniently extinct (so long, *Basilosaurus*), so we can ignore this fuzziness, but it is often there (and paleontologists are confronted with it more often). Think about the details of a species changing from one discrete state to another, even for something like a seemingly perfectly discrete character like a base changing from an A to a T. At first this is present in just a single individual (for a multicellular diploid, on one DNA strand in one cell in the germ line). Even if under selection, it will take generations to sweep through to fixation: during that time, what is “the” state of the species? It is even harder to discretize characters like woodiness (how much wood is required?), eyes (when does a fish population evolving in a cave finally “lose” its eyes?), biogeography (how finely do you divide the range: by continent? biome? state?), and so forth.

As for many decisions, this comes back to the biological hypotheses being tested and the size of the study. For example, one question could be does a complex trait like wings ever re-evolve once lost? Whiting et al. (2003) examined this in stick insects: some species have wings in both sexes, some in one only, and some lack wings in both sexes. If the question hinges on whether loss of wing genes in a species prevents re-evolution, then as long as one sex in a species has wings the species should be coded as having wings. If the question hinges on the effect of loss of wings on ability to settle new areas, it could be that having either sex lack wings is enough to prevent effective colonization, and thus a species with only one sex with wings should be coded as being wingless. If the study system is large enough to have sufficient power, one could code this as a four state character, instead: **A**: both males and females have wings; **B**: males have, females lack wings; **C**: males lack, females have wings; and **D**: both males and females lack wings.

One way to deal with this is to gather discrete data as finely divided as reasonable and then aggregate. For example, in the stick insect example, code it as a four state character as above and then, depending on the biological question, group them. If the question is whether wings can reappear after being entirely lost, for

example, one would group **A**, **B**, and **C** as having wings (in at least some members of the species, so the genes remain under selection for functionality) and **D** as wingless, but for the dispersal question one could lump **B**, **C**, and **D**, leaving **A** as the other state, or even lump **B** and **C** only.

| males | females | four_states | question_1 | question_2a | question_2b |
|----------|----------|-------------|------------|-------------|-------------|
| wings | wings | A | 1 | 1 | 2 |
| wings | wingless | B | 0 | 1 | 1 |
| wingless | wings | C | 0 | 1 | 1 |
| wingless | wingless | D | 0 | 0 | 0 |

But, let's assume we can discretize traits and carry on. The simplest discretization is binary: 0 or 1, often absence or presence (but could be yellow or blue, etc.). Most models are like our commonly used DNA models: continuous time with discrete changes, using the same rates throughout. It is like a model for when an autonomous car will have an accident: assuming the car works perfectly (gives a whole new meaning to "blue screen of death") there's still a chance that at some point a human is going to run into it. There's a per hour chance of an accident: let's assume in each hour there's a 0.03% chance of our autonomous car having an accident (very roughly based on Google's experience, assuming a 40 MPH average speed). So the probability of having no accident in the first hour of driving is 99.97%; the probability of having no accidents in the first 40 hours of driving is $99.97\% ^{40} = 98.8\%$. The number of accidents is Poisson-distributed; the wait time between accidents is exponentially distributed. This is the model commonly used in phylogenetics for discrete traits, though sometimes with more complexity: one could move (with some rate) between two different rates, as in a covarion model, for example. A very different model is Felsenstein's threshold model, which we will discuss in a few weeks. For now, though, just envision models with a fixed rate of change between states as long as other characters don't change; it's possible, though, that the state of other characters do affect these rates (which is what correlation tests investigate). For example, the probability of switching from clawed feet to flippers for forelimbs is probably much higher for species that live in water than on land.

(btw, note the spelling here: having one, two or eight eyes is a *discrete* trait: individually separate and distinct. Forming an enclosed bower for hidden mating is a *discreet* trait. The former is generally far more biologically relevant)

Chapter 9

Diversification

Chapter 10

SSE methods

Chapter 11

RAxML

Chapter 12

Gene Tree Species Tree

Chapter 13

Dating

Chapter 14

Build a method

Chapter 15

Placeholder

Chapter 16

Appendix

Bibliography

Whiting, M. F., Bradler, S., and Maxwell, T. (2003). Loss and recovery of wings in stick insects. *Nature*, 421(6920):264–267.