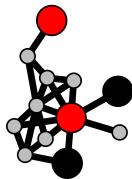


# The Statistical Dangers of Between-Species Comparisons

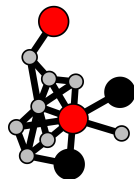
Jake L. Weissman

July 9, 2020



# The Comparative Framework

- Instead of pop-gen which looks at within-population genetic variation, we are interested in assessing evolutionary patterns across species
- Different types of questions
  - Do sets of traits tend to evolve together?
  - Do certain traits/innovations lead to bursts of diversification?
  - What was the phenotype of some extinct ancestor?

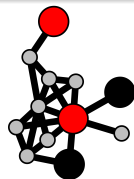


# Correlated Traits: An Example

- I am interested in looking for traits that predict the pathogenicity of a species of bacteria, using trait data for a large set of species
  - e.g., is genome size associated with pathogenicity?

## A (Bad) Solution:

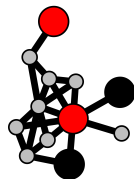
Calculate the correlation between genome size and pathogenicity



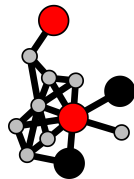
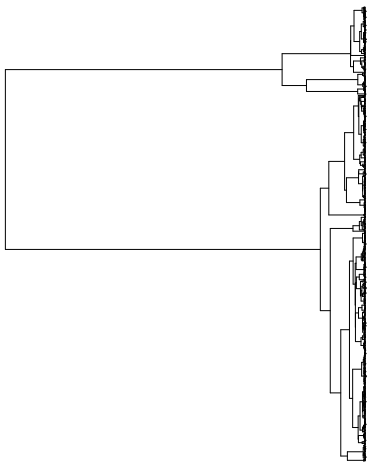
# The Problem: Hierarchical Structure

(aka phylogenetically structured data)

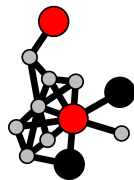
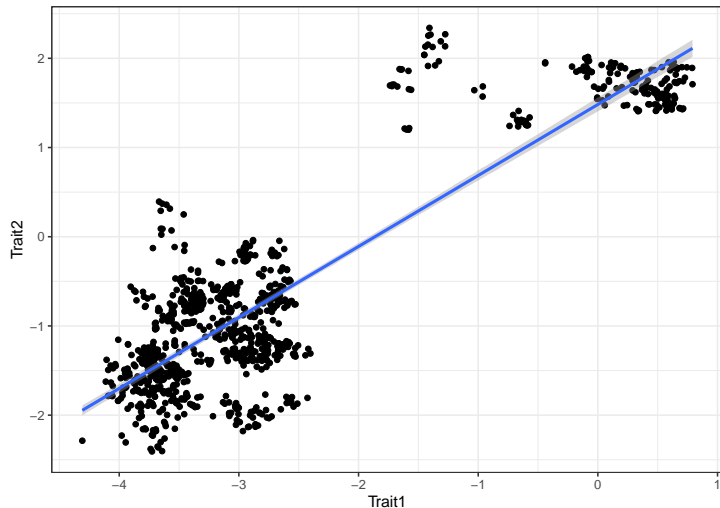
Your datapoints are not independent! (go read Felsenstien 1985)



# Imagine a Tree



# Uncorrelated Trait Evolution Simulated on Our Tree



# The Moral of This (Phylogenetically Structured) Story

- Phylogeny matters!
- If you don't take dependency structures in your data into account you are likely to get spurious (but very significant) results
- See R tutorial for methods on how to correct for phylogenetic structure (PGLS)

