Integrative Biology 200   "PRINCIPLES OF PHYLOGENETICS"   Spring 2014
University of California, Berkeley   Kipling Will- 21 Feb

Molecular dating and the use of fossils

By definition the extant tips of the tree are the same age, i.e. the *time* that has passed from the common ancestral node is the same for all extant OTUs. However, we also know that ultrametric trees are extremely rare for real data, so the amount of change along the branches differs across the tree most of the time.

Why there is no universal molecular clock:
generation time
population size
DNA repair efficiency
metabolic rates
selective pressures

A short digression: Missing Data and partitioning data. Both of these have the potential to significantly impact branch length estimates and the phylogeny and so are concerns for dating.

Molecular dating methods merge temporal information from sources outside of the primary phylogenetic data in order to provide time calibration for a phylogeny with branch lengths, corrected for rate variation. These may be strict or relaxed clock methods. There are a number of popular uses of molecular dating.

Biogeography- Establishing the origin of a fauna or testing dispersal and vicariance scenarios requires having an absolute time scale.

Age of the common ancestor- Correlating the origin of clades to climate or other events requires having an absolute time scale.

Diversification rates- Variation of evolutionary rates within or between clades or trees (co-diversification) require relative or absolute time scales.

Populations- Looking at the emergence and propagation of viruses and other disease pathogens, invasive species or other population-level questions may require relative or absolute time scales.

Calibration. Setting the age of one or several nodes in the tree using a point estimate, mean value or probability distribution allows for estimation of the age of other, uncalibrated nodes.

Sources of calibration:

Geology. Calibrations can be based on the distribution of the members of the clade given splitting events such as vicariance caused by orogeny or rifting when the event is well known and has an established date. Obviously shouldn't be used when the biogeographic history of the group is the question being addressed.

Secondary calibration. Dates from prior dating analyses or using rates for the same gene based on a published estimate can be done. However, this multiplies the error and uncertainty.

Fossils. Probably the most common and best method of calibration is to use a fossil. There are, however, <u>several sources of uncertainty</u>.

-Phylogenetic placement.
        Intuitive. Often based on overall similarity. Without an analysis there is no way to know the clade indicated really is a relative.
        Apomorphy. Optimization may cause a problem and outside of a analysis convergence can't be excluded.
        Phylogenetic
                Direct
                Indirect
                Reconciling molecular and morphological trees
                Multiple placements
-Determining the fossil's age.
        What method and how precise?
        Has there been a review of the dating of the fossil or strata
        Has the geologic scale changed?
        Correlated vs. direct dating
-The number, age and placement of multiple fossils.
-Identification.
        Must be a single physical specimen
        Incorrect
        Over-determined or over-identified
        Different taxonomic concepts

In the best possible case you will have multiple fossils to establish hard minimum dates, i.e. *the youngest possible age of the oldest known fossil*, that are 1. based on real documented specimens, 2. have a confirmed identification, 3. consistently and explicitly dated using a standard geologic time scale, and 4. placed in the phylogeny with the same set of OTUs as those with molecular data.

<u>Autocorrolation or non-autocorrolated</u>

Autocorrolated- Descendant branches draw a rate from a distribution with a mean given by the ancestral branch.

Non-autocorrolated- rates for each branch are drawn independently from an identical distribution.

Prior trees or simultaneous estimation.

Many dating studies estimate the tree using a time-independent method followed by the estimation of divergence dates using a molecular clock (strict or relaxed).
-Independence of models and unlinking assumptions
-relatively computationally easy and fast

Others do a simultaneous estimation of the tree, its branch lengths and time estimates (e.g. Beast).
-confidence intervals
-phylogenetic uncertainty, given that there is information for estimating dates of divergence, this could be used to give better estimates of phylogeny
-flexible clock models
-Computationally difficult

Some examples:
Local clock models (PAML, QDate) – clock-like within a given clade, but variable between clades.

Non-parametric rate smoothing (r8s) - simultaneously estimate unknown divergence times and smooth the rapidity of rate change along lineages

Ad hoc heuristic rate smoothing (PAML)

Penalized likelihood (r8s) Bayesian approach and penalized likelihood both have in common that they smooth or minimize rate variation over evolutionary time by means of an autocorrelated process.

Bayesian relaxed-clock methods (multidivtime, MrBayes, PhyBayes, BEAST).