Molecular dating and the use of fossils

Molecular dating methods are used to produce chronograms that are either relative and show the temporal relationships within a phylogeny or 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 to produce an absolute chronogram.

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 so relative chronograms can be examined for node depth to see the relative timing of diversification.

Branch lengths are the distance between OTUs and/or nodes that reflect both character state changes and time. To separate state changes and time we can use strict or relaxed clock methods. However, we know that truly ultrametric trees are rare for real data as the amount of change along the branches across the tree differs most of the time.

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

But data may be clock-like and the parameters for a clock model (nodal depth, n-1) are fewer than for a clock model (branch lengths, 2n-3). So we can test if the additional parameters result in a model that significantly better explains the data.

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.

Calibration. To make an absolute chronogram we set 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, several sources of uncertainty.

-*Phylogenetic placement.*

Intuitive, narrative. 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 an 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.

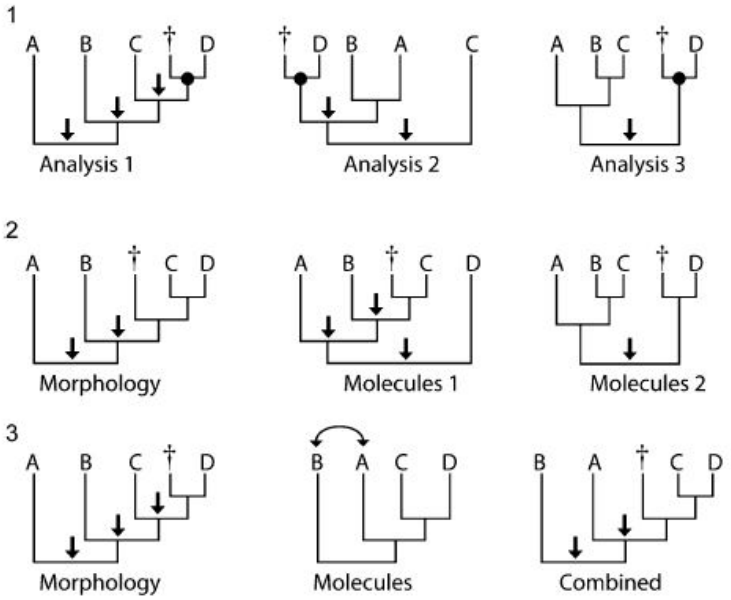

FIGURE 1. Example 1: A fossil (†) with unambiguous synapomorphies can be assigned to a specific lineage (D) with confidence. Regardless of the topology, the fossil will track the extant lineage and serve as a candidate calibration for all nodes above which it is nested. Example 2: Competing phylogenetic hypotheses from different data sets can change the position of fossil calibrations. In the morphological analysis, a fossil is found to be closely related to lineages C and D. Two arrows show the nodes that the fossil could calibrate. A molecular study with a different topology separates lineages C and D, making the placement of the fossil ambiguous. If the fossil is closely related to C, then it could calibrate three nodes. If the fossil is closely related to D, then it is a candidate calibration for just one node. Example 3: Changes to outgroup topology can change the polarization of morphological characters and placement of fossils. In the morphological analysis, a fossil (†) is placed in the C + D clade, sister to D. A molecular analysis changes the relationships of the outgroups (A and B). In a combined analysis, the morphological characters for the C + D clade are polarized in a different way and so using the fossil to calibrate clade C + D would be inappropriate.

Figure from Parham et al. 2012. Sys. Bio. 61.

Relaxed clock models

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

There are a number of popular uses.

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.