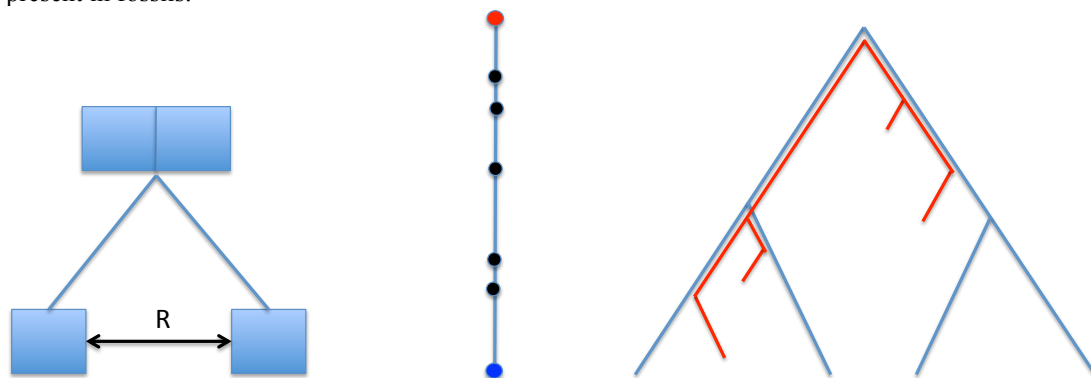


Using the molecular clock to test biogeographical hypotheses

Finn Borchsenius, Anders Barfod and Jotun Hein 1.9.11

Recent developments in molecular divergence time estimation offers means for comparing phylogenetic and biogeographic hypotheses. Following the general acceptance of Wegener's continental drift theory in the late 1960s vicariance became accepted as the predominant mechanism in historical biogeography due to its ability to offer a unifying explanation for parallel disjunctions in unrelated groups. A number of studies using molecular clock have shown that known geological processes believed to have caused vicariance do not fit expected patterns and/or expected timing of cladogenesis. Instead alternative explanations involving interplate dispersal routes during warmer periods of the Earth's history or long distance dispersal by sea currents or rafting on floating islands following prevailing wind directions as unifying explanations for repeated trans-oceanic dispersals. Crucial to the validity of these explanations are, however, that nodes can be reliably dated using molecular clock estimates.

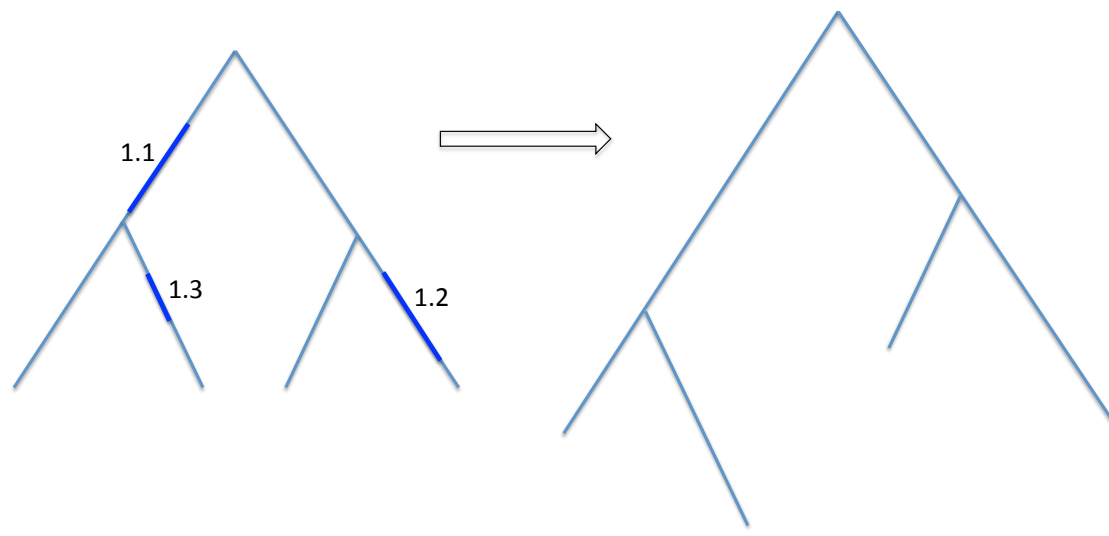
Stochastic Models underlying testing of biogeographical hypotheses have a series of facets compared to standard phylogenies. *Firstly*, due to the emphasis on dating and the simultaneous unrealism of the molecular clock, it is necessary to use some parametrized departure from the strict molecular clock, such as the fluctuating rates model of Thorne, Kishino and Painter (1999). *Secondly*, the geographical location of the present species [such as different continents] it is advantageous to work with a model with 2 or more areas that split at some point in geological time (vicariance) and allow for limited migrations after the split (dispersal). The introduction of area has the interesting consequence that splits in the species tree caused by vicariance may result in identical sister taxa. *Thirdly*, the molecular clock is calibrated with data from fossils of known age (incl. error estimates) and location. This latter is complicated by the fact that fossils cannot be referred to the inner branches of phylogenies constructed from extant molecular evidence. The method by Tavaré et al. (1998) estimates time intervals for fossil under an assumption of uniform sampling. Due to the fragmented nature of plant fossils, however, it is possible that remains of closely related species erroneously are considered conspecific. Although the fossil tree and tree for the extant species would overlap, both would have disjoint parts. This problem is reminiscent of the problems treated in Forsberg et al. (2005). Modeling fossil evolution is not straightforward. The continuous quantities or shape of fossils can be modeled by Brownian Motion or diffusion processes and used phylogenetically such as in Felsenstein (2004) or Thompson (1975). Phylogenetic use shape analysis [such as Dryden and Mardia, 1999] would be an optional choice, but has only been explored a little. Another approach would be to reconstruct the evolution of selected combinations of morphological characters on molecular trees of extant species. This would allow for minimum age estimates of the origin of particular combinations of characters present in fossils.



Left. A continent split up T years ago and forced a set of simultaneous speciation events. Possibly there has been low migration between the two continents since with rate R . Middle: a set of dated fossils from one species is used to estimate the interval in which the species existed [starting in red ball and ending in blue ball]. This can be used to put time restrictions on nodes of a phylogeny. Right: a phylogeny can be inferred using both fossil [ancient] and molecular data.

Phylogenies of selected groups of South American palms will be available as exemplar groups to

the project. South America is particularly interesting from a biogeographic point of view because of the recent upheaval of the Andes and the complex fluvio-dynamics of the Amazon region caused by alluvial deposits and microtectonic movements.



Left: a real time phylogeny with epochs of increased evolutionary rates. This defines a non-clock tree relating the leaves, that will have a limited and parametrized departure from clockness and thus used to make statements about the true real time tree. The fluctuations for evolutionary rates can be introduced in a variety of ways. Each edge having its own rate is an easy, but unbiological way to introduce this [why should rate changes coincide with speciation events?]. Alternatively epochs can be defined randomly as in Thorne et al. (1998), but could also be more gradual.

Methods: It would be natural to use either a likelihood or Bayesian method based on an explicit model or sequence/fossil evolution. Some of this can be done using existing phylogenetic packages such as Rev. BAYES [Huelsenbeck and Ronquist, 2001], but a deeper analysis would necessitate implementation of new methods.

Extensions: The use of genomic data would benefit from the use of statistical alignment methods [such as Novak et al. (2008)] and stochastic models of genomic rearrangements [Miklos et al. (2005)].

Project:

- Read basic literature
- Formulate explicit likelihood model for combined continent/fossil/genome data
- Analyze provided data set

References:

Thorne, Kishino & Painter (1998) "Estimating the rate of evolution of the rate of molecular evolution" *Mol Biol Evol* 15 (12): 1647-1657

Simon Tavaré, Charles R. Marshall, Oliver Will, Christophe Soligo & Robert D. Martin (2002) Using the fossil record to estimate the age of the last common ancestor of extant primates *Nature* 416, 726-729

Dryden, I. and K. Mardia (1998) "Shape Analysis" Wiley

Forsberg, Drummond and Hein (2005) Tree measures and the number of segregating sites in time-structured population samples. *BMC Genetics*, 6 Article: 35

Huelsenbeck JP, Ronquist F. (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*. 17(8):754-5.

Nomenjanahary Alexia Raharinirina, Liutauras Rusaitis, Henry Jackson (2011) "Modelling leaf shape evolution" Oxford summerschool in computational biology report

E. Thompson (1975) *Human Evolutionary Trees* CUP

Felsenstein, J. (2004) *Inferring Phylogenies. Chapter 23* Sinauer Associates, Sunderland, Massachusetts

Novák, Á., Miklós, I., Lyngsoe, R., Hein, J. (2008) StatAlign: An Extendable Software Package for Joint Bayesian Estimation of Alignments and Evolutionary Trees. *Bioinformatics*, 24(20), 2403-2404

Miklós, I., Itzész, P. & Hein, J. (2005) ParIS Genome Rearrangement Server. *Bioinformatics*, 21, 817-820