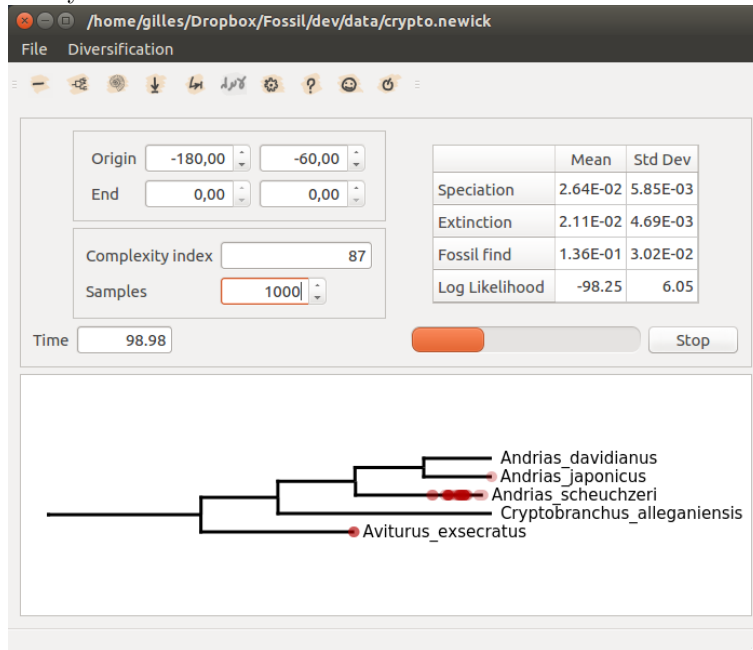# Diversification help

Diversification estimates diversification (speciation and extinction, separately) and fossilization rates from phylogenetic tree shapes and fossil times (geological ages) only. The software has tree design and editing capabilities via drag-and-drop and the contextual menu of the display zone. Fossils may be added and edited in the same way.


Setting trees and fossils

Diversification starts with a tree made of a single leaf, which can be edited to fit your purpose. Initiating a new tree is done by selecting the menu item *New tree* or the tool ![tool]. Alternatively, a tree in Newick format, or a tree with fossil information, can be loaded by using the menu item *Open tree* or the tool ![tool]. Fossils may next be set manually or imported from a CSV file by using the menu item *Import fossils* or the tool ![tool]. Before estimating the rates, there is important information to provide:

- the *Origin* which sets the time range for the start of diversification (which should be at least as old as the base node of the tree, but may be a bit older),

- the *End* which sets the time range of the end of diversification.

Editing trees and fossils

The tree topology may be modified by dragging branches from a point to another. A right click on a branch of the tree opens the contextual menu:

> Edit
> Remove
> Add leaf
> Add Fossil
> Edit Fossils
> Save subtree as
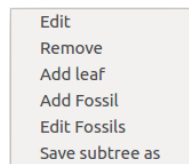
The menu item *Edit* opens the edit dialog of the leaf or of the internal node pending from the selected branch. For a leaf, it is

Edit leaf

Name

> Andrias scheuchzeri

Status
- ○ Contemporary
- ● Extinct
- ○ Not considered after:
  > 0,00   0,00

Cancel   OK

The edit dialog allows you to set the the name of the leaf and, more important, its status. *Contemporary* and *Extinct* has to be understood as usual. *Not considered after* means that the tree likelihood is computed as if the lineage was observable at, but with a fate unknown after, the range of dates below (or rather a date drawn in this range). The edit dialog of a node only allows you to set its name.

The menu items *Remove*, *Add leaf* and *Add fossil* just do what they say. *Add leaf* opens the preceding dialog and *Add fossil* asks for a range of fossil dates.

The menu item *Edit fossils* opens the dialog

Edit fossils

Bound inf	Bound sup
-17,85	-17,75
-17,80	-17,75
-17,80	-17,70
-17,40	-17,20
-17,00	-16,70
-14,90	-14,50

Add   Remove

Cancel   OK

Each line of the table displays the dating interval of a fossil. You can edit an entry of the table by double-clicking on it, and use the eponymous buttons to add and remove fossil intervals of times.

The menu item *Save subtree as* saves the subtree pending from the selected branch in Newick format (with fossils).

Saving trees and estimates

The menu item *Save* and the tool ![save tool] save the current tree in a Newick format enhanced to include fossils, origin and end time information, which are stored as "NHX-like" tags in the comments of the nodes. The menu item *Export* and tool ![export tool] allows you to export the tree picture in various formats (PDF, PNG, SVG and PSTricks), the tree itself in Newick format and the rate estimates as tables in CSV format or as text reports. The contextual menu of the display zone has an option to save subtrees in Newick format.

Estimating the rates

Computation time

The complexity of the likelihood computation, thus its running time, depends not only on the tree size but also on the tree shape and on the fossil density and position (fossils on internal branches help reduce computation time, compared to fossils on terminal branches). The *Complexity index* gives a hint of the computation time, which also depends on your computer. The field is not editable and is updated each time the tree changes. The figure below plots the computation time of the likehood (with Intel® Xeon(R) 2.80GHz) versus the complexity index of various random trees.
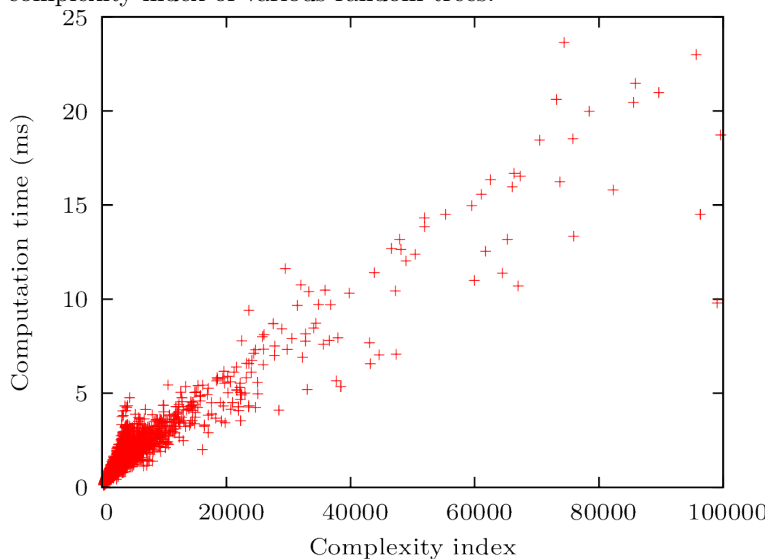
Optimization settings

The maximum likelihood estimation is performed by using numerical optimizations. In short, the optimization process seeks for local optimum of the likelihood of the tree in the space of rates by starting from random points.

The optimization settings are modified via the menu item *Options* or the tool ![options tool]. It opens the dialog:

The fields *Speciation* (resp. *Extinction*, *Fossil*) set the interval in which the start speciation rates (resp. extinction rates, fossilization rates) are uniformly drawn. Note that it does not mean that the optimizer seeks solution only in these intervals. But they are important since starting for a point too far from the optimal one increases the chances of not converging to the global optimum, thus of getting inacurrate estimates. The field *Trials* sets the number of times the optimization process is launched, each time with a new random starting point. The fields *Tolerance* and *Max iteration* are for the stopping conditions of the optimization. The process stops either when two successive points differ less than the tolerance or when the number of iterations is greater than the max. Tolerance influences the precision of the estimates.

The optimizer settings are crucial. We advise you to try several settings and check both the *Std Dev* column and the *Log Likelihood* line the until you get satisfying results. A possible strategy is to start with large intervals for the starting points of the rates (maybe with a greater number of trials but less sampling times - see below), then reduce their sizes according to the estimates obtained. The aim is to get standard deviations of the estimates as small as possible and likelihoods as high as possible.

Sampling times

Our likelihood computation does not directly deal with intervals of possible times for fossil ages, the origin and/or the end of the diversification process. Before estimating the rates, *Diversification* uniformly samples exact times from the intervals. The field *Samples* set the number of time samplings which are performed. Each of them gives an estimate of the rates. The mean rate estimates and their standard deviations are displayed in the top-right table.

Rates estimation

The estimation of the diversification and fossil find rates is launched by the item *Compute* of the Diversification menu or the tool . The mean rate estimates and their standard deviations are updated each time an estimation from a time sampling is complete. The process may be stopped by clicking on the *Stop* button right to the progress bar (the computation is actually stopped at the end of the current threads, which may take a while).

Please address feedbacks to `gilles.didier@univ-amu.fr`

Gilles Didier
Institut de Mathématiques de Marseille (UMR 7373)
Site Sud, Campus de Luminy, Case 907
13288 MARSEILLE Cedex 9