Estimation of dN and dS considering different substitution categories

We present a tutorial to estimate dN and dS for different categories of substitutions, either overall, or depending on the change in GC content.

`S` (resp. `W`) stands for `G` or `C` (resp. `A` or `T`). And, for example, `S` → `W` is any substitution from `S` to `W`, and `dS X S` → `W` is any synonymous such substitution.

Maximum likelihood estimate of model and branch lengths

The first step is to infer the best-fit parameters for a specific model, root and tree from the data by maximum likelihood. Here, we use the YN98 model.

```
bppml param=base.bpp
```

`TENT.dnd` is the starting tree for the optimization procedure. The only important feature in this tree is the topology, which is not optimized. The branches are numbered in the order they appear in the file, for example here: `(((0,(1,2)3)4,(5,6)7)8,9);` (for user declaration, use Nhx format (see bpp-suite manual)). The optimized tree and model are respectively in files `tree_ml.dnd` and `model_ml.params`[1].

Estimates of dN and dS

Then, to compute dN and dS on each branch, we use `mapnh` with the optimal tree and model obtained previously.

We get one tree file per type of substitution, with branch lengths replaced by the branch specific estimates.

```
mapnh param=map_dNdS.bpp
```

The dN and dS counts on all branches are respectively in tree files `TENT.counts_dN.dnd` and `TENT.counts_dS.dnd`.

An example of usage in R is in file `manip.R`. In this file, we use the library `ape` to handle trees. Note that the node numbers are different between `bio++` and `ape`.

Category specific estimates of dN and dS

With command

```
mapnh param=map_dNdS_GC.bpp
```

[1] In the configuration file we set the tolerance to stop the optimization process rather high, `optimization.tolerance = 10`, to shorten the computation time. The user can check that with lower values the estimates of the parameters are not much different.

we compute dN and dS for each category `S` → `W`, `S` → `S`, `W` → `W`, `W` → `S`.

The dN and dS counts for the different categories on all branches are found in the tree files, such as
`TENT.counts_dN_X_W->S.dnd` and `TENT.counts_dS_X_W->S.dnd`.