

Update for *Analysis of Phylogenetics and Evolution with R*

This document gives some updates to *Analysis of Phylogenetics and Evolution with R* since the text of the book was finalized (February 2006). It points to the new functions and options that have been done since then. These updates are grouped according to the chapters of the book. The package concerned is indicated in the margin.

Chapter 1

Some recent packages available on CRAN are worth mentioning.

- **geiger** by Luke Harmon and colleagues provides several functions for “Running macroevolutionary simulation, and estimating parameters related to diversification from comparative phylogenetic data.”
- **laser** by Dan Rabosky provides several functions for the likelihood analysis of speciation and extinction rates from phylogenies, as well as simulation of phylogenetic trees under several constant rates models.
- **paleoTS** by Gene Hunt provides functions for the analysis of paleontological evolutionary times-series. Though the book focuses on phylogenies of recent species, this package deals with some of the issues addressed there but with fossil data.
- **phangorn** by Klaus Schliep provides several methods for the estimation of phylogenies including ML, NJ, UNJ, a parsimony-based method for DNA sequences, and Hadamard method. It also includes several utilities to convert among some formats to code DNA sequences, simulation of sequences on a tree, and an implementation of the Shimodeira–Hasegawa test.
- **phybase** by Liang Liu provides basic functions for phylogenetic analysis. It includes the function `maxtree` that estimates a species tree from multiple gene trees.
- **PhySim** by Jason Weir and Dolph Schluter “contains functions to simulate phylogenetic trees under a virth death model. Functions are provided to model a lag-time to speciation and extract sister species ages.”
- **picante** by Steve Kembel and colleagues provides tools for integrating phylogenetics and ecology, namely “Phylocom integration, community analyses, null-models, traits and evolution in R”.

Two other packages on BioConductor are also worth mentioning.

- **Biostrings** by Saikat DebRoy provides “string objects representing biological sequences, and matching algorithms”.
- **GeneR** by L. Cottret and colleagues provides functions for “manipulating nucleotidic sequences (Embl, Fasta, GenBank)”.

Chapter 3

- The internal structure of the class "phylo" has been changed substantially: see `?read.tree` for details. This requires less memory, and most operations are now much faster. The trees are now printed in a compact way on the console (see below). ape
- `read.tree` can read trees with a single branch which are correctly represented as an object of class "phylo": ape

```
> read.tree(text = "(a);")
```

```
Phylogenetic tree with 1 tips and 1 internal nodes.
```

```
Tip labels:
```

```
[1] "a"
```

```
Rooted; no branch lengths.
```

- Lists of trees have now the class "multiPhylo", and their manipulation, plotting, and storage have been improved. ape
- The function `root` has a new option, `resolve.root`, to avoid a trichotomy at the root. ape
- The functions `pandit` and `treebase` have been replaced by a single function `dbtrees`; its first argument `db` specifies the tree database: apTreeshape

```
dbtrees("pandit", ...)  
dbtrees("treebase", ...)
```

So Fig. 3.1 can be drawn with:

```
plot(dbtrees("pandit", 2), font = 1)
```

- DNA sequences are now internally stored in a binary format; several functions are provided to manipulate them efficiently (see `?DNABin` for details). There is also now a function `as.alignment` to convert to the "alignment" class of `seqinr`. ape
- The new function `del.gaps` removes insertion gaps in DNA sequences. ape
- The new function `makeLabel` helps to manage labels of trees or of DNA sequences (e.g., truncating them, or making them unique). ape
- The new function `makeNodeLabel` creates or modifies node labels in a flexible way. ape
- The new function `is.monophyletic` tests the monophyly of a group. ape

Chapter 4

- `plot.phylo` has a new argument `tip.color` to specify the colors of the tips. It is used in the same way than `font`. See Fig. 4.20 for an example; a colored version could be: ape

```
plot(tr, tip.color = c(rep("blue", 15), rep("red", 3)))
```

Another new argument of this function is `edge.lty` to draw branches as dotted, dashed, etc., lines. It is used in the same way than `edge.color`.

- Trees can now be plotted as circular trees with the option `type = "fan"`. ape
- There is a `plot` method for lists of trees (objects of class "multiPhylo"): it calls `plot.phylo` internally and is documented on the same help page. The option `layout` controls the number of trees plotted at the same time (one by default). ape
- `nodelabels` and `tiplabels` have a new argument `pie` to represent proportions, with any number of categories, as piecharts. The use of the option `thermo` has been improved: there is now no limitation on the number of categories. ape
- Giving the new structure of the class "phylo", the appearance of Fig. 4.7 would have changed. Also, Fig. 4.8 would have been drawn specifying 4 and 5 as node numbers (commands on p. 73).
- The new function `edgelabels` adds labels in the middle of the edges of a plotted tree. The options are the same than for `nodelabels` and `tiplabels`. ape
- The new function `identify.phylo` helps to find information about a tree by clicking on it when plotted. ape
- The new function `cophyloplot` can plot directly two trees face-to-face, greatly simplifying the procedure described on page 87. Two other new functions, `subtrees` and `subtreeplot`, help to explore large phylogenies. ape

Chapter 5

- The function `dist.dna` has three new models: Barry–Hartigan (1987), paralinear, and LogDet. ape
- Matrix exponential can now be computed with the new function `matexpo`. ape
- The function `dist.alignment` should be included in Table 5.1: it computes a simple dissimilarity matrix from aligned nucleotidic or amino acid sequences. seqinr
- The new function `sh.test` computes the Shimodaira–Hasegawa test. ape
- Three new functions, `bionj`, `fastme.ols`, and `fastme.bal`, perform phylogeny estimation with the BIONJ (Gascuel, 1997, *Mol. Biol. Evol.* 14:685) and minimum evolution methods (Desper & Gascuel, 2002, *J. Comput. Biol.*, 9:687). ape

- The new function `chronMPL` estimates the node ages of a tree using the mean path lengths method (Britton *et al.*, 2002, *Mol. Phyl. Evol.*, 24:58). ape
- The new functions `CADM.global` and `CADM.post`, contributed by Pierre Legendre, test the congruence among several distance matrices. ape

Chapter 6

- The function `drop1.compar.gee` (used as, e.g., `drop1(m)`) tests for single effects in GEE-based comparative method. This may be useful with discrete predictors with more than two categories where printing the results gives only the *t*-tests of the parameters. A warning message is printed if there is not enough degrees of freedom. ape
- `ace` has a new argument `ip` that specifies the initial values used by the ML estimation with discrete characters (see the examples in `?ace`). This function now returns a matrix giving the indices of the estimated rates when analysing discrete characters. ape
- There are four new ‘method’ functions to be used with the results of `ace`: `logLik`, `deviance`, `AIC`, and `anova`. The latter is particularly useful to compute the likelihood ratio tests. For instance, the test in the middle of page 174 would be performed with: ape

```
anova(syl.er, syl.sym)
```

- The function `maxlik.betasplit` computes the maximum likelihood estimate of the parameter β of Aldous’s β -splitting model. apTreeshape
- The function `shift.test` tests for a change in diversification rate on a given node of a tree according to a procedure derived by Moore, Chan & Donoghue (2004). apTreeshape
- The new function `yule.time` fits a user-defined time-dependent Yule model by maximum likelihood. ape

September 28, 2009