APE: Analyses of Phylogenetics and Evolution in R language

Emmanuel Paradis[1,*], Julien Claude[1] and Korbinian Strimmer[2]

[1]Laboratoire de Paléontologie, Paléobiologie and Phylogénie, Institut des Sciences de l'Évolution, Université Montpellier II, F-34095 Montpellier cédex 05, France and [2]Department of Statistics, University of Munich, Ludwigstrasse 33, D-80539 Munich, Germany

ABSTRACT

Summary: Analysis of Phylogenetics and Evolution (APE) is a package written in the R language for use in molecular evolution and phylogenetics. APE provides both utility functions for reading and writing data and manipulating phylogenetic trees, as well as several advanced methods for phylogenetic and evolutionary analysis (e.g. comparative and population genetic methods). APE takes advantage of the many R functions for statistics and graphics, and also provides a flexible framework for developing and implementing further statistical methods for the analysis of evolutionary processes.

Availability: The program is free and available from the official R package archive at http://cran.r-project.org/src/contrib/PACKAGES.html#ape. APE is licensed under the GNU General Public License.

Contact: paradis@isem.univ-montp2.fr

Phylogenetic analysis, in its broad sense, covers a very wide range of methods from computing evolutionary distances, reconstructing gene trees, estimating divergence dates, to the analysis of comparative data, estimation of evolutionary rates and analysis of diversification. All these diverse tasks have one particular aspect in common: they rely heavily on computational statistics.

The R system, a free platform-independent open-source analysis environment, has recently emerged as the *de facto* standard for statistical computing and graphics (Ihaka and Gentleman, 1996). One advantage of R is that it can be easily tailored to a particular application area by writing specialized packages. In particular, the usefulness of R in bioinformatics has already been impressively demonstrated in the analysis of gene expression data (http://www.bioconductor.org).

Analysis of Phylogenetics and Evolution (APE) is the first joint effort to utilize the power of R also in the analysis of phylogenetic and evolutionary data. APE focuses on statistical analyses using phylogenetic and genealogical trees as input.

In Version 1.1, APE provides functions for reading, writing, plotting and manipulating phylogenetic trees, analyses of comparative data in a phylogenetic framework, analysis of diversification, computing distances from allelic and nucleotide data, reading nucleotide sequences and several other tools, such as Mantel's test, computation of minimum spanning tree or estimation of population genetics parameters. Table 1 gives an overview of the functions currently available in APE. Note that some of the methods (e.g. comparative method, skyline plot, etc.) have previously been available only in specialized softwares. External tree reconstruction programs (such as PHYLIP) can be called from R through standard shell commands.

One strength of R is that it is straightforward to obtain publication-quality graphical output, particularly with its PostScript device. For instance, the plotting function of phylogenies in APE handles colors, line thickness, font, spacing of labels, which can be defined separately for each branch, so that three different variables can be represented on a single phylogeny plot. APE also produces complex population genetics plots, such as the generalized skyline plot (Strimmer and Pybus, 2001), with a single command.

APE, like any R package, is command-line driven. The functions are called by the user, possibly with arguments and options. Any session using APE in R starts with the command

```
library(ape)
```

which makes the functions of APE available in the R environment. The list of these functions can be displayed with the command

```
library(help = ape)
```

which displays their names with a brief description. An evolutionary tree saved on the disk in the text file `tree1.txt` in the standard Newick parenthetic format can then be read by

```
tree1 <- read.tree('tree1.txt')
```

This stores the phylogenetic tree is in an *object* named `tree1` of class '`phylo`'. The information stored in this object

---

*To whom correspondence should be addressed.

Table 1. Special functions available in APE 1.1

Application	Available commands
Input/output	`read.dna`, `write.dna`, `read.nexus`, `write.nexus`, `read.tree`, `write.tree`, `read.GenBank`
Graphics	`add.scale.bar`, `plot.mst`, `plot.phylo`, `plot.skyline`, `lines.skyline`, `ltt.plot`
Tree manipulation	`bind.tree`, `drop.tip`, `is.binary.tree`, `is.ultrametric`
Comparative method	`compar.gee`, `compar.lynch`, `pic`, `vcv.phylo`
Diversification	`birthdeath`, `cherry`, `diversi.gof`, `diversi.time`, `gamma.stat`
Population genetics	`branching.times`, `coalescent.intervals`, `collapsed.intervals`, `find.skyline.epsilon`, `heterozygosity`, `skylineplot`, `skyline`, `theta.h`, `theta.k`, `theta.s`
Molecular dating	`chronogram`, `ratogram`, `NPRS.criterion`
Miscellaneous	`all.equal.phylo`, `balance`, `base.freq`, `dist.dna`, `dist.gene`, `dist.phylo`, `GC.content`, `klastorin`, `mantel.test`, `mst`, `summary.phylo`
Data sets	`bird.families`, `bird.orders`, `hivtree`, `landplants`, `opsin`, `woodmouse`, `xenarthra`

Detailed information about each function can be accessed with the online help [e.g. `help(mantel.test)`].

(e.g. branch lengths) can be inspected by typing `tree1` and graphical output in form of a cladogram can be obtained by executing

```
plot(tree1)
```

which actually calls the function `plot.phylo` of APE to draw the phylogenetic tree `tree1` [due to the object-oriented nature of R the command `plot(x)` may give a completely different result depending on the class of `x`]. The tree is plotted, by default, on a graphical window, but can be exported in various file formats depending on the operating system.

In addition to this trivial example, the representation of a phylogenetic tree in an object-oriented structure results in straightforward manipulation of the phylogenetic data for various computations used in evolutionary analyses. Currently implemented in APE are approaches, such as phylogenetically independent contrasts (Felsenstein, 1985; Harvey and Pagel, 1991), fitting birth–death models (Nee *et al.*, 1994; Pybus and Harvey, 2000), population-genetic analysis (Nee *et al.*, 1995; Strimmer and Pybus, 2001), non-parametric smoothing of evolutionary rates (Sanderson, 1997) and estimation

of groups of genes in phylogenetic trees using Klastorin's method (Misawa and Tajima, 2000). Furthermore, distance-based clustering methods as implemented in the R function `hclust` can be used by APE using functions converting to and from objects of class 'phylo' and 'hclust'.

All R functions available in APE (Table 1) are documented in the R hypertext format and information regarding their use can be retrieved by applying the `help` command, e.g.

```
help(read.tree)
```

The classes and methods in APE (like `phylo`) can also easily be further extended to include other functionalities, for instance to annotate phylogenetic trees. Thus, APE is not only a data analysis package, it is also an environment for developing and implementing new methods. Furthermore, programs written in C, C++ or Fortran77 can be linked and called from R. This is particularly useful for computer intensive calculations.

ACKNOWLEDGEMENTS

REFERENCES

Felsenstein,J. (1985) Phylogenies and the comparative methods. *Am. Nat.*, **125**, 1–15.

Harvey,P.H. and Pagel,M.D. (1991) *The Comparative Method in Evolutionary Biology*. Oxford University Press, Oxford.

Ihaka,R. and Gentleman,R. (1996) R: a language for data analysis and graphics. *J. Comput. Graph. Statist.*, **5**, 299–314.

Misawa,K. and Tajima,F. (2000) A simple method for classifying genes and a bootstrap test for classifications. *Mol. Biol. Evol.*, **17**, 1879–1884.

Nee,S., Holmes,E.C., Rambaut,A. and Harvey,P.H. (1995) Inferring population history from molecular phylogenies. *Phil. Trans. R. Soc. Lond. B*, **349**, 25–31.

Nee,S., May,R.M. and Harvey,P.H. (1994) The reconstructed evolutionary process. *Phil. Trans. R. Soc. Lond. B*, **344**, 305–311.

Pybus,O.G. and Harvey,P.H. (2000) Testing macro-evolutionary models using incomplete molecular phylogenies. *Proc. R. Soc. Lond B*, **267**, 2267–2272.

Sanderson,M.J. (1997) A nonparametric approach to estimating divergence times in the absence of rate constancy. *Mol. Biol. Evol.*, **14**, 1218–1231.

Strimmer,K. and Pybus,O.G. (2001) Exploring the demographic history of a sample of DNA sequences using the generalized skyline plot. *Mol. Biol. Evol.*, **18**, 2298–2305.