

---

## Preface to the Second Edition

Some kinds of problems may yield to sophisticated statistical technique; others will not. The goal of empirical research is—or should be—to increase our understanding of the phenomena, rather than displaying our mastery of technique.

—Freedman [91]

The first edition of *APER* was published five years ago. The subject of the book has grown in importance and audience during this time, so this second edition is certainly timely. Moreover, the changes in `ape` and several related packages have been so important in the last few years, that, for some time, I have felt this new edition was required.

Though I acknowledge that the first edition of *APER* was quite appreciated by researchers, lecturers and students, it was not out of criticism. With this second edition, I have tried to improve on all possible points and issues as much as I was able to do. At the same time, I had to think over on the progress of the approach I presented five years ago. A difficult task was to review the development of the many new packages contributing to phylogenetics and evolutionary biology. It was my choice to focus on some of these packages with the aim to provide to the reader a consistent set of “techniques” to address a wide range of issues in phylogenetics and macroevolution. Even restricting our interest to these two fields, a number of packages are not considered in this book: this is a choice I fully assume. Besides, I decided to not tackle two related fields: population genetics and paleobiology. Both have seen critical developments in the form of several packages. A useful treatment of these would require two new books, so I have limited myself here to only mention some data structures in Chapter 3.

The general structure of this second edition is very similar to its predecessor. Chapters 1–4 focus on the “technique” of R and several specialized packages for phylogenetics and evolutionary biology with the aim to give to the readers the tools to manipulate, summarize, and display their data. Chapters 5 on estimation of phylogeny, and 6 on analysis of macroevolution,

are more concerned with “understanding the phenomena” of evolution. Compared to the first edition, these two chapters have been much expanded and emphasized. Though I have tried to go deeper in the inferential and biological processes, a complete treatment of these questions would require several volumes, and is therefore limited here to the essential points. The new Chapter 7 covers the simulation of evolutionary data, a topic which, I believe, deserves to be treated with more rigor than done until now. Chapter 8 concludes the book on the development of phylogenetic computational methods in R. This chapter has been expanded using essentially my experience over the last few years. Because the volume of the book has been nearly doubled, the number of the case studies has been reduced.

At the time of completing this second edition, a number of packages have been pre-released or are under development. These will, in the years to come, increase dramatically the quantity of software available to evolutionary biologists to analyze their data. Even the packages described in details in this book continue their development. For instance, **ape** will soon include most published methods to estimate phylogenies from incomplete distance matrices thanks to the effort of Andrei-Alin Popescu. Some other packages, such as **ade4**, **diversitree**, **phangorn**, or **seqinr**, have provided a range of tools and methods to many evolutionists for some time, and are under continuous development.

An interesting phenomenon emerged recently in the scientific literature: many authors provide R code as supplementary information to their articles. Though we can be worried about the proliferation of sources of R programs, this is very good news for our speciality as it marks a significant progress towards a wider acceptance of the concept of repeatable research in data analysis. This, and the above, testify that using R as a computational tool for phylogenetics and evolutionary biology is meant to last.

Several colleagues kindly read some parts of the manuscript of this second edition: Christophe Douady, Susan Holmes, Nicolas Hubert, Anthony Ives, and Errol Rowe. I am grateful to Thibaut Jombart for reading several chapters and for the stimulating discussions, and to Klaus Schliep for our many discussions. Many thanks also to my colleagues in Nescent for organizing the Hackathon in December 2007. Thank you to Régis Hocdé for helping me to obtain the web site [ape.mpl.ird.fr](http://ape.mpl.ird.fr), and the repository for the source of **ape**; thank you also to the IRD staff for maintaining them over these years. Writing a book is a tremendous experience. Having the chance to correct and amend it is another unique experience. I thank warmly John Kimmel and Marc Strauss for giving me this opportunity. I am sincerely thankful to a long list of users, colleagues, and friends for their interest and enthusiasm.

I am grateful to Laure and Sinta for their constant support and patience.

---

## Preface to the First Edition

As a result, the inference of phylogenies often seems divorced from any connection to other methods of analysis of scientific data.

—Felsenstein [78]

Once calculation became easy, the statistician's energies could be devoted to understanding his or her dataset.

—Venables & Ripley [307]

The study of the evolution of life on Earth stands as one of the most complex fields in science. It involves observations from very different sources, and has implications far beyond the domain of basic science. It is concerned with processes occurring on very long time spans, and we now know that it is also important for our daily lives as shown by the rapid evolution of many pathogens.

As a field ecologist, for a long time I was remotely interested in phylogenetics and other approaches to evolution. Most of the work I accomplished during my doctoral studies involved field studies of small mammals and estimation of demographic parameters. Things changed in 1996 when my interest was attracted by the question of the effect of demographic parameters on bird diversification. This was a new issue for me, so I searched for relevant data analysis methods, but I failed to find exactly what I needed. I started to conduct my own research on this problem to propose some, at least partial, solutions. This work made me realize that this kind of research critically depends on the available software, and it was clear to me that what was offered to phylogeneticists at this time was inappropriate.

I first read about R in 1998 while I was working in England: I first tried it on my computer in early 1999 after I got a position in France. I quickly thought that R seemed to be the computing system that is needed for developing phylogenetic methods: versatile, flexible, powerful, with great graphical possibilities, and free.

When I first presented the idea to develop programs written in R for phylogenetic analyses in 2001, the reactions from my colleagues were mixed with enthusiasm and scepticism. The perspective of creating a single environment for phylogenetic analysis was clearly exciting, but some concerns were expressed about the computing performance of R which, it was argued, could not match those of traditional phylogenetic programs. Another criticism was that biologists would be discouraged from using a program with a command-line interface. The first version of the R package **ape** was eventually released in August 2002. The reactions from some colleagues showed me that related projects were undertaken elsewhere.

The progress accomplished has been much more than I expected, and the perspectives are far reaching. Writing a book on phylogenetics with R is an opportunity to bring together pieces of information from various sources, programs, and packages, as well as discussing a few ideas.

I realize that the scope of the book is large, and the treatment may seem superficial in some places, but it was important to treat the present topics in a concise manner. It was not possible to explore all the potentialities now offered by R and its packages written for phylogenetic analysis. Similarly, I tried to explain the underlying concepts of the methods, sometimes illustrated with R codes, but I meant to keep it short as well.

I must first thank the “R community” of developers and users from whom I learned much about R through numerous exchanges on the Internet: this definitely helped me to find my way and envision the development of **ape**. Julien Claude has shared the venture of developing programs in R and contributing to **ape** since he was a doctoral student. A great thank you to those who contributed some codes to **ape**: Korbinian Strimmer, Gangolf Jobb, Rainer Opgen-Rhein, Julien Dutheil, Yvonnick Noël, and Ben Bolker. I must emphasize that all these authors should have full credit for their contributions. I am grateful to Olivier François and Michael Blum for showing me the possibilities of their package **apTreeshape**.

Several colleagues kindly read some parts of the manuscript: Lounès Chikki, Julien Claude, Jean Lobry, Jean-François Renno, Christophe Thébaud, Fabienne Thomarat, and several colleagues who chose to remain anonymous. Thanks to all of them! Special thanks to Susan Holmes for encouragement and some critical comments. Thank you to Elizabeth Purdom and Julien Dutheil for discussions about **ape** and R programming. I am sincerely thankful to John Kimmel at Springer for the opportunity to write this book, and for managing all practical aspects of this project. Finally, many thanks to Diane Sahadeo for handling my manuscript to make it an actual book.

<http://www.springer.com/978-1-4614-1742-2>

Analysis of Phylogenetics and Evolution with R  
Paradis, E.

2012, XIV, 386 p. 89 illus., Softcover

ISBN: 978-1-4614-1742-2