

Penetrating the black box of phylogenetic analysis

Lena C. Hileman

University of Kansas, 1200 Sunnyside Avenue, Lawrence, KS 66049, USA.

lhileman@ku.edu

doi:10.1242/dev.038182

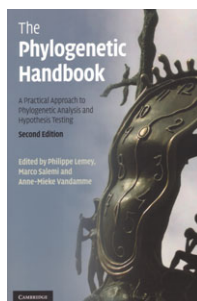
The Phylogenetic Handbook: A Practical Approach to Phylogenetic Analysis and Hypothesis Testing, Second Edition

Edited by Philippe Lemey, Marco Salemi and Anne-Mieke Vandamme

Cambridge University Press (2009) 723 pages
ISBN 978-0-521-73071-6 (paperback), 978-0-521-87710-7 (hardback)
£29.99/\$59 (paperback), £70/\$130 (hardback)

Developmental genetic studies frequently include some form of comparative analysis; perhaps an amino acid alignment to identify conserved domains encoded within your gene of interest, or a gene tree to assess orthology or paralogy relationships with other members of a gene family. Developmental evolutionary (evo-devo) studies usually also include these sorts of analysis, but might also include assessments of character evolution across phylogenies, or patterns of molecular evolution across gene trees. Therefore, for many developmental or developmental evolutionary biologists, being able to understand and apply phylogenetic methods is an essential skill. However, the reality is that many people come to the task as novices, and that it is generally very difficult to keep pace with the ever-expanding fields of phylogenetic and molecular evolutionary analyses. Wouldn't it be great if there were a user-friendly handbook available that kept pace with this field? Editors Philippe Lemey, Marco Salemi and Anne-Mieke Vandamme provide just that in the second edition of *The Phylogenetic Handbook: A Practical Approach to Phylogenetic Analysis and Hypothesis Testing*.

In this edition of *The Phylogenetic Handbook*, the editors have taken what, in my opinion, was a decent handbook for phylogenetic analysis, namely the first edition, reorganized it and have incorporated new content, thus making the second edition a great handbook. Nearly without exception, the chapters are presented at a level that will guide both the novice and the more experienced user through what are often considered to be the 'black boxes' of phylogenetic and molecular evolutionary



analyses. In order to facilitate an understanding of both theory and practice, each chapter is presented in two parts. First, the background and the basic theory for a particular type of data analysis (e.g. multiple sequence alignment, phylogenetic inference using maximum likelihood methods, natural selection and adaptation of molecular sequences) are presented. Each theory section is then followed by a practice section, which walks the reader through phylogenetic or molecular evolutionary analyses using examples of software packages and data sets. These are easily accessed and downloaded through links on the book's website (<http://www.thephylogenetic handbook.org>).

Despite the fact that this is an edited volume, which could easily lead to inconsistencies between chapters in terms of the clarity and the accessibility to the non-expert reader, the theory sections of the handbook are, nearly without exception, highly informative and should be accessible to a novice user of phylogenetic methods. The practice sections of each chapter are, however, what truly make this an eminently useful and user-friendly handbook. In each practice section, readers receive step-by-step instructions that allow them to take sample data sets through a number of phylogenetic and molecular evolutionary analyses, using selected software packages. Importantly,

Wouldn't it be great if there were a user-friendly handbook available that kept pace with this field? [The] editors provide just that

the reader is instructed and warned about critical model and parameter choices that must be made when executing analyses, and usually the reader is referred back to the theory section to ensure that informed, as opposed to 'black box', choices are made. A further plus is that some chapters

also provide information on how to use newer tree-viewing software that aids the transition from phylogenetic analysis to publication-quality tree images.

The new, much expanded second edition of *The Phylogenetic Handbook* contains 21 chapters organized into eight thematic areas. By comparison, the first edition contained 15 chapters that were presented without higher-level organization. The themes range from data preparation and the use of sequence databases; to multiple sequence alignment and phylogenetic inference (including distance, parsimony, maximum likelihood and Bayesian criteria); to the testing of models, including models of molecular evolution, molecular clock, and tree topologies; to the use of codon-based models of evolution for detecting natural selection/adaptation of molecular sequences.

The additional content included in the second edition reflects the major advances in the field that have been made over the past ten years. Of most importance for developmental and developmental evolutionary biologists will be the new and/or expanded chapters that cover multiple sequence alignments, Bayesian phylogenetic analysis, the selection of models of evolution for phylogenetic analyses, the testing of alternative tree topologies, and the detection of a history of natural selection/adaptation using codon-based models of evolution. Even if you own a copy of the first edition, you should still consider replacing it with the second edition because it covers important advances in the field that have taken place since the publication of the first edition.

In addition to the book itself, the accompanying website (www.thephylogenetic handbook.org) provides some additional useful information. For example, Table 3.1 from the text can be found on the website, which lists software used for multiple sequence alignment, together with references and links to websites where the software can be downloaded. In fact, for each of the software packages discussed in the handbook, the website provides links to specific sites where it can be downloaded. All the sample data sets provided for the practice sections of the handbook are also easily downloaded from the website, and the website contains an errata page where changes to the availability of software (i.e. new URLs) that have occurred since the publication of the handbook can be found. The website will presumably be kept updated, which should allow this edition of the handbook to stay on top of further progress in the field for some time.

Will *The Phylogenetic Handbook* eliminate all the ‘blood, sweat and tears’ that can accompany phylogenetic and molecular evolutionary analyses for the ‘non-card carrying’ phylogeneticist? Of course not. I am confident, however, that the use of this book as a guide can significantly reduce the trauma frequently associated with such analyses. To get a sense of its utility, I went ahead and ran some data sets of my own through a handful of the analyses that are outlined in the practice sections of the book. I found that stumbling blocks remain when it comes to properly installing or executing software, or to properly formatting data sets so that they run smoothly through various software packages. In addition, the software packages that are highlighted in the practice sections of *The Phylogenetic Handbook* might simply not be the particular software packages that you would like to use. There are, for example, many software options for multiple sequence alignment or for phylogenetic analysis using the maximum likelihood criterion, but only a few are presented in the corresponding practice sections of the handbook. You will have to navigate the installation and the execution of these other software packages on your own. This might seem daunting, but at least the theory section of the handbook can equip you with the knowledge needed to make informed decisions about your methodology (and keep in mind that, although this new edition of *The Phylogenetic Handbook* is highly informative and user-friendly, nothing compares to befriending your local phylogenetics guru).

Of course, there are limits to what can be included in the handbook itself. To get around the drawback that only select software is presented in the practice sections, it would be excellent to see additional content added to the website. It would be especially helpful to include on the website the same type of step-by-step practice exercises that are presented in the handbook for additional, commonly used software packages, such as DIALIGN, RAXML, GARLI, and PAML. Such additions to the website would, however, only further the usefulness of this already very handy handbook. I am convinced that its background, theory and practice sections are invaluable for guiding readers through the proper design and execution of phylogenetic and molecular evolutionary analyses, and recommend *The Phylogenetic Handbook* whole-heartedly to anyone whose work might venture in this direction.

Models for the future

Sebastian Shimeld

Department of Zoology, University of Oxford, Oxford
OX1 3PS, UK.

Sebastian.shimeld@zoo.ox.ac.uk

doi:10.1242/dev.039271

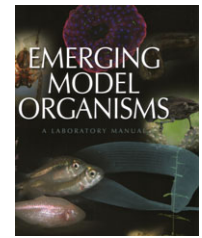
Emerging Model Organisms: A Laboratory Manual, Vol. 1

Edited by Richard R. Behringer, Alexander
D. Johnson and Robert E. Krumlauf

Cold Spring Harbor Laboratory Press (2009) 592 pages
ISBN 978-087969872-0
\$89 (paperback)

Any biologist who has worked with a genuine ‘model organism’, be it plant or animal, mould or microbe, knows the powerful insight into biology that can be provided by such species. In part this insight derives from the intrinsic characters that underlie their initial adoption as models, such as a rapid generation time and ease of husbandry. But their usefulness also depends on the extent to which their intrinsic traits have been built upon; the sheer depth of knowledge of how they work; the development of extensive resources in the form of genes, mutants and transgenic lines; and the establishment of incisive experimental methods. When combined, these allow a fidelity of experimentation that is, at least to me, breathtaking. The arguments as to why we as a scientific community should invest so much in such organisms are overwhelming. With such a focus of resources we can go further, faster; we can then take the details of what we learn and apply them to situations of economic and therapeutic need. We may also seek to establish the underlying principles of how such organisms work, and then apply these in a predictive manner across the tree of life. The success of these approaches should be apparent to all.

Why, then, should we be interested in taking into the laboratory other species, variously known as non-model organisms, new model organisms or, in the case of this book, emerging model organisms? One powerful argument is that every model system has its own unique evolutionary history and hence its own set of adaptations, specialisations and caveats. The workhorses of developmental biology, the fruitfly *Drosophila melanogaster* and the nematode *Caenorhabditis elegans*, are a case in point; immensely powerful in many ways, it has recently become clear that both species have lost a surprisingly large number of genes



that were present in the common ancestor of the bilaterians, and that are still found in our own genome (Kortschak et al., 2003; Miller and Ball, 2008; Putnam et al., 2007; Technau et al., 2005). New models will be needed to bring the power of invertebrate genetics to bear on the functions of these genes.

A second argument, and one that I suspect motivates many (including myself), is the drive to understand the evolutionary basis for the diversity of life. Studies restricted to a few species will never illuminate how enigmatic groups such as the cephalopods evolved from their snail-like molluscan ancestors, or how vertebrates evolved the complex cranial senses, jaws and brains associated with their transition to a predatory lifestyle. For these questions, we need to focus on species that diverged at informative points in evolution. For these reasons (and others), many researchers have been turning to a host of different species, some previously studied in other contexts and others essentially new to experimental

Will this information help a lab trying to set up a system that is new to them? The answer is undoubtedly, yes

biology. *Emerging Model Organisms* is aimed at helping such adoption by providing centralised collections of information on resources, tools and protocols for a wide diversity of relevant species.

The book is divided into 23 chapters, each describing a different species and each written by a research group with expertise in the area. The species chosen are mostly multicellular, reflecting a focus on developmental biology, and range from the social amoeba *Dictyostelium* (which has such a well-developed research community, resources and tools it could uncharitably be called a ‘well-emerged’ model organism) to the dogfish *Scyliorhinus*, in which little is currently possible experimentally beyond