

BOOK REVIEW

Cladistics: A Practical Primer on CD-ROM. By Peter Skelton and Andrew Smith, with accompanying booklet by Neale Monks. Cambridge University Press, 40 West 20th Street, New York, NY 10011-4211, ISBN: 0-521-52341-9, Price \$55.00

Cladistics, or phylogenetic systematics, is an important tool for every evolutionary biologist. Following Hennig's groundbreaking work (Hennig, 1966) the goal of modern taxonomists is to develop a natural system of classification, i.e. a system that is based on the organisms' shared ancestry. Yet many students find it very difficult to absorb the terminology of cladistics, and consequently to use it in their research. Peter Skelton's and Andrew Smith's CD, *Cladistics: a practical primer on CD-ROM* (Skelton and Smith, 2002), is a great place to start. It introduces the basic concepts through examples, and offers many interactive exercises along the way. It has an easy to navigate interface, and the material is divided into modules that can be listened to separately. The terms that occur in the text are hyperlinked to the glossary that pops up in a separate window upon clicking. The only minor deficiencies of the program interface are its inability to pause the narrative within a module, and to adjust the font size.

Step by step, the CD introduces cladistics terminology, the concept of homology, types of data used in systematics, how to build trees from morphological and molecular data, as well as some statistical measures to assess the reliability of the obtained tree and how to root it. The course culminates in a practical chapter that allows the reader to test the learned material by performing an actual phylogenetic analysis. The CD spends considerable amounts of time on tree topologies and their manipulation, which is one of the topics that novices to the field often struggle with, and on an overview of the principle of parsimony. The authors succeed in keeping things simple, and they avoid many potentially confusing and contentious details. Unfortunately, as it often happens through oversimplification, some inaccuracies are introduced. For example, the authors define 'network' as a dichotomously branching unrooted tree. Although every tree is a network, the reverse is not true, which makes the provided definition incorrect. The definition of 'long-branch attraction' as a tendency of molecular sequences to accumulate base substitution through time such that the actual phylogenetic signal is lost is a definition for saturation with substitutions; the latter *might lead* to the long-branch

attraction in the tree reconstructions. However, the long-branch attraction should be rather defined as an artifact that tends to group long branches in the tree together. When the authors discuss the possible changes that can occur within the molecular sequences, such as deletions or insertions, they use an example of a deletion that occurred in one of the two homologous sequences. But with only two sequences it is impossible to distinguish between an insertion in one sequence and a deletion in the other sequence; without additional data both of these events appear equally parsimonious. Wrong conclusions may be drawn if the deletion is favored over the insertion. The definition of an indel as insertion or deletion would solve the problem.

Another shortcoming is that the authors ignore many aspects crucial for successful phylogenetic analyses based on molecular data: alignment algorithms, phylogenetic tree reconstruction methods other than parsimony and among site rate variation. The only considered alignment procedure is alignment by eye; global or local alignment algorithms are not discussed. Distance-based tree reconstruction methods are not mentioned at all, and maximum likelihood methods are mentioned once or twice, but never discussed. This is all the more confusing as the authors describe the most parsimonious tree as the most likely one.

One hallmark of cladistic analysis is polarization of characters, i.e. the rooting of a tree. It is usually difficult to polarize molecular characters; notable exceptions are large insertions like the non-homologous region in the vacuolar and archaeal type ATPases (Gogarten *et al.*, 1989). As a result phylogenies from molecular data usually are unrooted and need to be rooted by other means. For the methods to introduce polarity, the authors concentrate on the ontogenic criterion and on the use of outgroup, but ignore the rooting of trees using ancient duplication events. The latter method is important, since the ontogenic criterion is not always true or applicable (e.g., see Mabee, 1996), and appropriate species to be used as outgroups are not always available (Gogarten *et al.*, 1989; Mathews and Donoghue, 1999).

The accompanying booklet by Neale Monks provides a rather redundant repetition of the CD's contents; even the same illustrations are used. The CD would have benefited from a complementary booklet that illustrates the same topics through different examples and alternative explanations, and that provides suggestions for further reading. This would make it easier for the interested person to follow up on the details of the discussed topics. However, in its present form the booklet does not add much to the material on CD.

Overall, the CD is a valuable tool for people outside the field who wish to introduce themselves to the basic principles in cladistics, but not sufficient in detail for those who chose evolutionary biology as their major tool or profession. The CD fits excellent into the role of a complimentary study aid for an undergraduate introductory course in evolutionary biology.

Summary

Pros:

- Great for beginners: introduces cladistic terminology through comprehensive and interactive examples
- Easy to navigate interface
- Modular chapter structure
- Glossary that is linked throughout the lessons
- Pauses in the accompanying narrative facilitate comprehension of the material
- Interactive exercises at the ends of the sections

Cons:

- The booklet is redundant. The CD should be available without the booklet.
- No links to suggested further reading or the primary literature.
- No attention is given to analyses of protein sequences
- No attention is given to the alignment of homologous sequences
- As might be expected from a cladistic approach to classification, only parsimony based phylogenetic reconstruction is considered. No other methods are introduced; in particular, problems associated with substitutions occurring with different rates in different sites are not discussed.
- Some misleading oversimplifications

References

- Gogarten, J. P., Kibak, H. *et al.*: 1989, Evolution of the Vacuolar H⁺-ATPase: Implications for the Origin of Eukaryotes, *Proc. Natl. Acad. Sci. U.S.A.* **86**(17), 6661–5.
- Hennig, W.: 1966, *Phylogenetic Systematics*, University of Illinois Press, Urbana.
- Mabee, P. M.: 1996, Forum: Reassessing the Ontogenetic Criterion: A Response to Patterson, *Cladistics* **12**(2), 169–176.
- Mathews, S. and Donoghue, M. J.: 1999, The Root of Angiosperm Phylogeny Inferred from Duplicate Phytochrome Genes, *Science* **286**(5441), 947–50.
- Skelton, P. and Smith, A.: 2002, *Cladistics. A Practical Primer on CD-ROM*, Cambridge University Press, Cambridge.

Olga Zhaxybayeva*

J. Peter Gogarten

Department of Molecular and Cell Biology,

University of Connecticut,

Storrs, CT, 06269-3125

(* author for correspondence, e-mail: olga@carrot.mcb.uconn.edu)