Phylogenetics

Treevolution: visual analysis of phylogenetic trees

Rodrigo Santamaría∗,† and Roberto Therón†

Department of Computer Science and Automation, Pz. de Los Caídos S/N, 37008 Salamanca, Spain

Received on March 3, 2009; revised and accepted on May 20, 2009
Advance Access publication May 26, 2009

ABSTRACT

Summary: Treevolution is a tool for the representation and exploration of phylogenetic trees that facilitates visual analysis. There are several useful tools to visualize phylogenetic trees, but their level of interaction is usually low, especially in the case of radial representations. Highly interactive visualizations can improve the exploration and understanding of phylogenetic trees. Treevolution implements strategies to interact with phylogenetic trees in order to allow a more thorough analysis by users.

Availability: Treevolution is available at http://vis.usal.es/treevolution.
Additional figures, a user’s guide, a video demo and some examples are available at the same site.

Contact: rodr@usal.es

Supplementary information: Supplementary data are available at Bioinformatics online.

1 INTRODUCTION

Phylogenetics is the study of the evolutionary relationships among organisms. A phylogenetic tree conveys these relationships as a hierarchical tree structure, where each node represents an organism and it is connected to all its descendants and to its unique direct ancestor. These relationships come from well-established taxonomies, usually with hypothetical ancestors that sometimes are identified by paleontological discoveries, which help to estimate times of evolutionary branches (trees that represent evolutionary times as branch lengths are called chronograms).

Typically researchers use phylogenetic trees as a way to illustrate the results achieved in their works, but it is becoming a current need to interact with them in order to gain insight into the problem. There are several tools to visualize phylogenetic trees, such as iTOL (Letunic and Bork, 2007), a great generator of tree images, including features such as tree annotation and the display of horizontal gene transfers. However, its interaction options are limited to a basic zooming and pruning. On the other side, TaxonTree (Parr et al., 2004) is a tool for tree visualization with a higher degree of interaction (expansion/pruning of branches, history tracing, zoom, text search), but its linear node-link tree layout makes it difficult to properly visualize large trees.

When performing a visual analysis of phylogenetic trees, the presentation of results is the last step of an iterative process involving previous phases where visualization has a main role. The proposed tool helps to answer the questions that phylogenetic trees try to unravel, by means of interaction with the visualization.

2 METHODS

A tool implementing a visual analytics approach must provide the basis for an analytic discourse between the analyst and the information (Thomas and Cook, 2005). The main question a phylogenetic tree answers is: how are the organisms in the tree related? Specific instances of this question can generate hypotheses to be tested using an analytic discourse supported by visual exploration. Examples of it are does the expansion in diversity of mammals start with the disappearance of dinosaurs? or does Homo sapiens come from the largest subfamily of Eutheria?

Treevolution is a Java program that makes use of Processing (http://processing.org) to visualize phylogenetic trees in either Newick or PhyloXML formats as radial dendrograms. It is based on a previous generic tool to visualize hierarchical trees (Theron, 2006). Families are located at different sectors and, in the case of chronograms, rings and branch lengths convey periods of time (Fig. 1). Ancillary bar charts and linear dendrograms can also be visualized at the user’s request (Fig. 2d and e, respectively). Treevolution offers several methods to explore trees: radial or sector distortion, tree rotation, pruning, labeling, tracking of ancestors and descendants, text search, etc.

∗To whom correspondence should be addressed.
†The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.
Fig. 2. (a) At 98 Myrs, the main families of mammals are present: the four families of Placentalia (green, blue, yellow and pink), plus Marsupialia (orange) and Monotremata (a small family in red at the top of the circle). (b) After the Cretaceous/Tertiary boundary at 65.5 Myrs, with the extinction of dinosaurs, mammal subfamilies start to arise, but outburst does not occur, specially in Euarchontoglires [the green family in (a)], the largest family of Placentalia at the present, which includes Hominidae. (c) The true rise of mammals is delayed until about 50 Myrs, as revealed by the high number of colored subfamilies, specially in Euarchontoglires (left semicircle). (d) Bar chart representing the number of branches per million year. (e) Linear dendrogram for Felidae.

When analyzing large phylogenetic trees, it is interesting to identify families (groups of organisms with the same ancestor). In the case of chronograms, evolutionary ages (determined by branch length) act as threshold scales, permitting analysis of how families develop through time. Using rings as threshold limits, Treevolution performs visual clustering of families by means of color.

3 RESULTS

Figure 2 illustrates some static images of Treevolution. Figure 2a–c shows the phylogenetic tree for mammals as a radial dendrogram (each ring represents a million of years, Myrs). By means of visual clustering and tree exploration, Treevolution can help in the analytical discourse to answer, for example, the question proposed by Bininda-Emonds et al. (2007): how long after the Cretaceous-Tertiary boundary was the rise of mammals? The visual analysis confirms that the evolutionary burst of mammals was delayed about 10 Myrs after the C/T boundary. Other questions with answers intrinsic to the phylogenetic tree can be easily answered by the tool (see the web site http://vis.usal.es/treevolution, for other examples).

4 CONCLUSION

We have developed a tool to display and interact with phylogenetic trees according to the philosophy of visual analytics. Treevolution is specially suited for chronograms, helping to discuss and confirm hypotheses. We described here a small example of use, but Treevolution can be applied to different datasets and analysis needs.

Funding: MEC [project GRACCIE (CONSOLIDER-INGENIO, CSD-2007-00067)]; the JCyL [project GR34 and grant EDU/1453/2005].

REFERENCES


