

Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation

Ivica Letunic and Peer Bork*

EMBL, Meyerhofstrasse 1, 69117 Heidelberg, Germany

Associate Editor: Joaquin Dopazo

ABSTRACT

Summary: Interactive Tree Of Life (iTOL) is a web based tool for the display, manipulation and annotation of phylogenetic trees. Trees can be interactively pruned and re-rooted. Various types of data such as genome sizes or protein domain repertoires can be mapped onto the tree. Export to several bitmap and vector graphics formats is supported.

Availability: iTOL is available through WWW at <http://itol.embl.de>

Contact: bork@embl.de

1 INTRODUCTION

The amount and diversity of species with at least partial sequence information is rapidly increasing (Cole, et al., 2005; Tringe, et al., 2005) and the tree of life based on completely sequenced genomes is being constantly redrawn (e.g. (Ciccarelli, et al., 2006)). Phylogenetic trees represent a backbone for various other biological studies and it is therefore essential to have state-of-the-art tools for their display, customization and interpretation.

During the years, numerous tree visualization tools, both online and standalone, have been developed. Tools like TREEVIEW (Page, 1996), NJPLOT (Perriere and Gouy, 1996), ATV (Zmasek and Eddy, 2001), TaxonTree (Parr, et al., 2004) and others provide many useful features and various ways of tree visualization. However, with advances in research and data generation, there is a continuous need for the incorporation of novel features. Examples are i) visualization of mid-sized trees (100 – 2000 leafs) with branch lengths in acceptable resolutions, ii) possibility of annotating trees with various datasets and iii) display of horizontal gene transfers between branches. Here we describe a new tool, iTOL, which offers these features and incorporates a layout that was repeatedly requested by the community after its original publication in (Ciccarelli, et al., 2006).

2 FEATURES

iTOL's functions provide an easy way to manipulate and create customized graphical representations of phylogenetic trees in the standard 'New Hampshire' or Newick format.

In addition to the standard tree representation, trees can be displayed in a circular (radial) mode (Figure 1), which is particularly useful for the visualization of mid-sized trees (up to several thousand leaves). Circular trees can be rotated and displayed in user defined arc sizes.

Several functions are available which allow users to customize their tree displays in various ways. Branches can be pruned or collapsed and any node can be used to re-root the tree. Colors can be assigned to various leaf groups, simplifying the navigation around the tree. For trees whose leaf IDs are based on NCBI Taxonomy (Wheeler, et al., 2006), iTOL can automatically determine taxonomic classes of all internal nodes and assign proper scientific names to the leafs. Internal node labels are displayed in popup windows which appear when mouse pointer is positioned above the corresponding node (Figure 1).

2.1 Pruning and collapsing branches

Pruning is a process of selecting one or several branches from the original tree and creating a new, smaller tree. Sub-branches and individual leaves can be easily added or removed.

Branches whose detailed structure is not needed can be collapsed. Collapsed branches are displayed as triangles. Total branch lengths to the closest and the farthest leaf are used to calculate the lengths of the triangle's sides.

2.2 Displaying external data on a tree

iTOL can display several types of data directly on the tree (Figure 1). Up to five datasets can be uploaded with each tree. Datasets are contained in plain text files with each line corresponding to one leaf in the tree. Supported dataset types are: binary data, simple bars and stacked bars (multiple values associated with each leaf).

iTOL is the first visualization tool which supports the display of horizontal gene transfers (HGTs) annotated directly in the original Newick tree file. HGT information is encoded in the IDs of tree nodes. Therefore, in addition to standard leaf IDs, the tree must have unique IDs assigned to all internal nodes. Detailed explanation with example trees and datasets is available in the iTOL online help pages.

2.3 Exporting trees to other formats

Each tree display in iTOL (either default or customized) can be exported to several graphical formats, both bitmap and vector based. Currently supported formats are Scalable Vector Graphics (svg), Portable Network Graphics (png), Encapsulated Postscript (eps), Postscript (ps) and Portable Document Format (pdf). In addition, pruned trees can be exported as Newick plain text files.

3 IMPLEMENTATION

*To whom correspondence should be addressed.

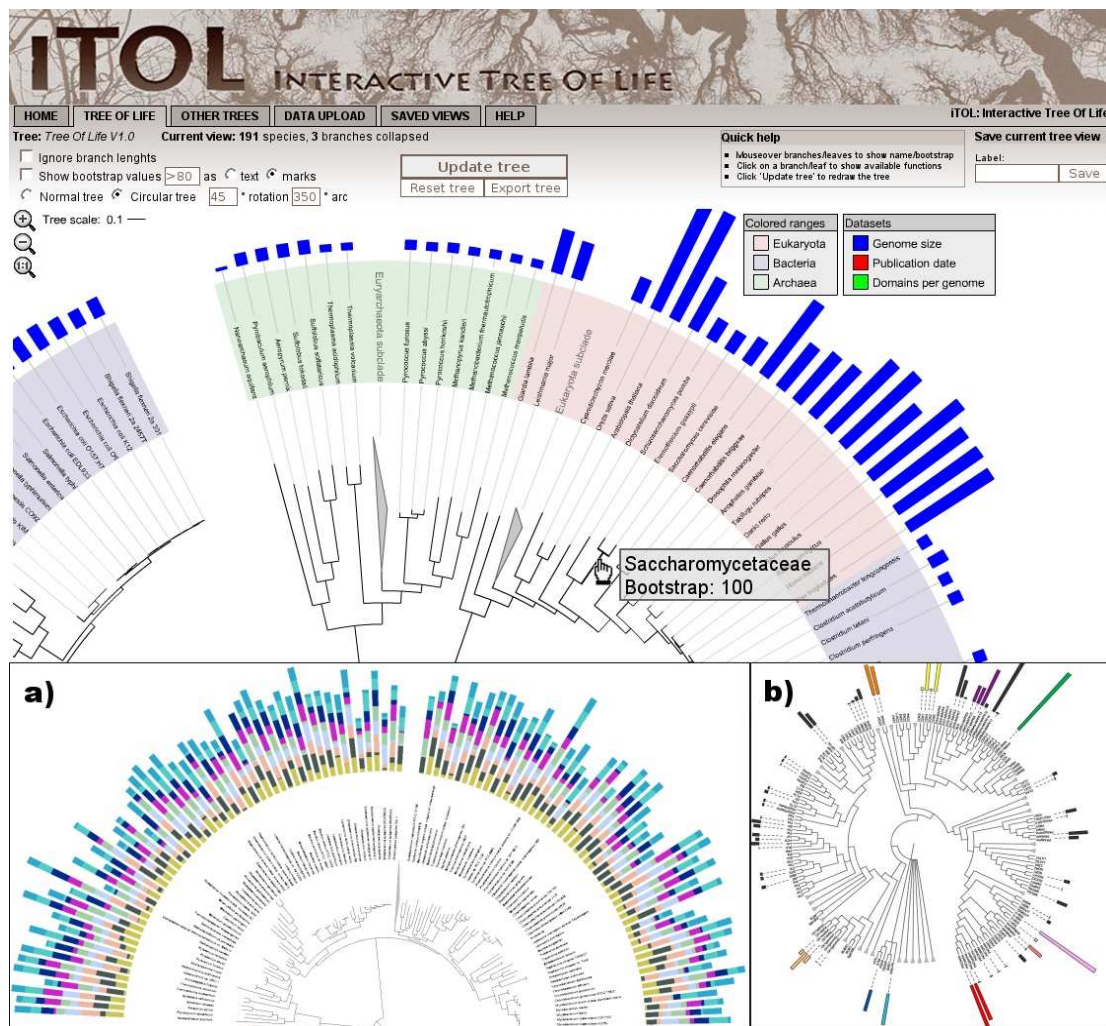


Fig. 1. iTOL's user interface and example trees. The interface with a tree of life (Ciccarelli, et al., 2006), annotated with genome sizes (blue bars), is shown. Several branches are collapsed and displayed as triangles. Popup windows give detailed information on branches/leaves, such as bootstrap values and taxonomical categories. **a)** A 200 species tree displayed in a 180° arc with a stacked bar dataset. **b)** A protein kinases tree displayed without branch length information and annotated with predicted site counts for various families.

iTOL is a WWW based tool, accessible using any modern web browser. It is coded in Shockwave Flash and JavaScript. The Flash object which displays the tree and provides interactivity is created dynamically by the server using a set of Perl scripts and the Ming library (<http://ming.sourceforge.net>). Flash plug-in version 7 or higher is required for full functionality. Tree export is done through an intermediary SVG (scalable vector graphics) file, which is converted to other formats using the Inkscape graphics package (<http://www.inkscape.org>). User uploaded trees and data are stored locally on the server and available for at least one year after original submission.

ACKNOWLEDGEMENTS

We would like to thank Christopher Creevey, Sean Hooper and other members of the Bork group for helpful suggestions and discussions.

REFERENCES

- Ciccarelli, F.D., et al. (2006) Toward automatic reconstruction of a highly resolved tree of life, *Science*, **311**, 1283-1287.
- Cole, J.R., et al. (2005) The Ribosomal Database Project (RDP-II): sequences and tools for high-throughput rRNA analysis, *Nucleic Acids Res.* **33**, D294-296.
- Page, R.D. (1996) TreeView: an application to display phylogenetic trees on personal computers, *Comput Appl Biosci*, **12**, 357-358.
- Parr, C.S., et al. (2004) Visualizations for taxonomic and phylogenetic trees, *Bioinformatics*, **20**, 2997-3004.
- Perriere, G. and Gouy, M. (1996) WWW-query: an on-line retrieval system for biological sequence banks, *Biochimie*, **78**, 364-369.
- Tringe, S.G., et al. (2005) Comparative metagenomics of microbial communities, *Science*, **308**, 554-557.
- Wheeler, D.L., et al. (2006) Database resources of the National Center for Biotechnology Information, *Nucleic Acids Res.* **34**, D173-180.
- Zmasek, C.M. and Eddy, S.R. (2001) ATV: display and manipulation of annotated phylogenetic trees, *Bioinformatics*, **17**, 383-384.