

## Phylogenetics

**Radié: visualizing taxon properties and parsimonious mappings using a radial phylogenetic tree**Jacqueline Whalley<sup>1</sup>, Stephen Brooks<sup>2</sup> and Robert G. Beiko<sup>2,\*</sup><sup>1</sup>School of Computing and Mathematical Sciences, Auckland University of Technology, Private Bag 92006, Auckland 1142, New Zealand and <sup>2</sup>Faculty of Computer Science, Dalhousie University, 6050 University Avenue, Halifax, Nova Scotia, Canada B3H 1W5

Received on November 16, 2008; revised and accepted on January 16, 2009

Advance Access publication February 4, 2009

Associate Editor: Martin Bishop

**ABSTRACT**

**Summary:** A phenotypic character has a convex mapping to a given phylogenetic tree if each character state can be assigned a single point of origin in the tree. However, phenomena such as convergent evolution and lateral genetic transfer can lead to intermingling of character states and a consequent non-convex mapping. The phylogenetic heterogeneity of different characters can identify subsets of states that are non-randomly associated or which may have been transferred from one lineage to another. We have developed Radié, an interactive 3D Java application for mapping character states to the leaves and internal edges of a radial phylogenetic tree. In Radié each state of a given character is associated with a unique color, and internal edges with many descendant character states can be represented in a number of different ways to illustrate the diversity within each group.

**Availability:** Radié is freely available for download at <http://kiwi.cs.dal.ca/~beiko/software-and-data/radie>; source code is available upon request from the authors.

**Contact:** [beiko@cs.dal.ca](mailto:beiko@cs.dal.ca)

**Supplementary information:** Supplementary data are available at *Bioinformatics* online.

**1 INTRODUCTION**

Character-based phylogenetic analyses often rely on a parsimony approach that aims to minimize the number of character state changes on a tree. The rooting of the eukaryotic tree based on a fusion of two conserved genes (Stechmann and Cavalier-Smith, 2002) is an example of such a minimization. Indeed, prior to the advent of small-subunit RNA sequencing, the classification of prokaryotes was based on the few phenotypic traits that could be discerned, including Gram stain, cell shape and metabolic pathways (Sapp, 2005). However, many of the relationships suggested by phenotypic similarity have now been refuted by whole-genome and whole-proteome analysis, with lateral genetic transfer (LGT) and convergent evolution as important mechanisms driving the patchy distribution of character states throughout the purported tree of species (Koonin and Wolf, 2008).

With these and other processes in mind, we can assess the extent to which character states are convex with respect to a tree, i.e.

can be mapped to a common point of origin that is exclusive of other character states. Stable traits that are rarely transferred, such as membrane lipid composition and gene fusion events should be convex or nearly so on a tree: such traits may have many underlying genes that cannot be transferred en bloc, or induce unacceptable fitness costs in most organisms. Conversely, traits such as antibiotic resistance that are frequently transferred, and traits such as genomic G+C content that can converge, are more likely to exhibit patchy distributions on a reference tree. The degree of patchiness can be quantified with the parsimony score, which calculates the minimum number of state changes needed to explain the distribution of character states over a reference tree, and most parsimonious ancestral states can be computed using the Fitch algorithm (Fitch, 1971). Together, these measures can be used to assess the relative stability of different traits.

We have developed a radial tree viewer that can be used to map and quantify the distribution of different traits onto a phylogenetic tree. The Interactive Tree of Life server (IToL, Letunic and Bork, 2007) allows the mapping of traits to the leaves of a radial phylogenetic tree, but we include multicolor representations of the set of states subtended by internal branches, and methods to compute the parsimony score of traits and most parsimonious internal node states using the Fitch algorithm.

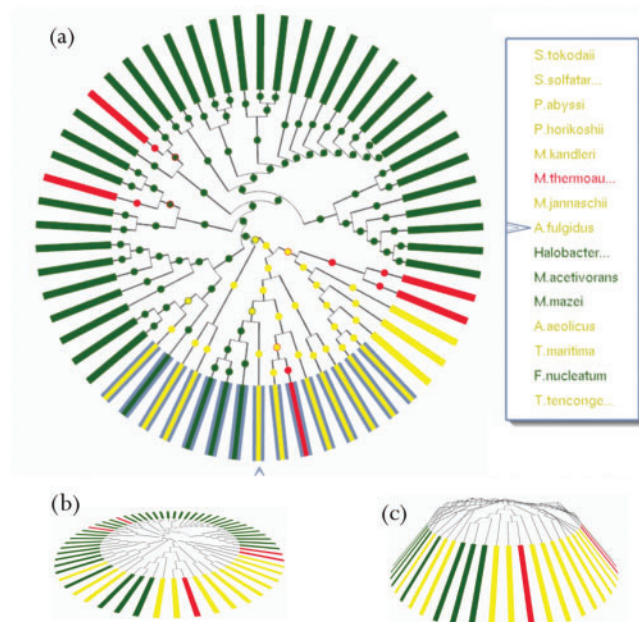
**2 IMPLEMENTATION AND DATA FORMAT**

Radié was developed using the Java OpenGL wrapper library JSR-231 (<http://jcp.org/en/jsr/detail?id=231>) which allows OpenGL to be used in the Java programming language and the use of Java Swing windows components. Radié reads in a NEXUS-formatted rooted tree which includes an extra character block to describe the leaf traits that may be mapped to the phylogenetic tree. The extra LEAF\_PROPERTIES block is designed to conform with the existing NEXUS standard (Maddison *et al.*, 1997), thus allowing existing tree files to be imported into Radié with only slight modifications. A full description and example of the LEAF\_PROPERTIES block can be found in the manual.

**3 DISPLAY AND ANALYSIS**

Phylogenetic trees may be viewed as a radial tree either as a 3D cone, in perspective or top-down (Fig. 1). In this figure, the

\*To whom correspondence should be addressed.



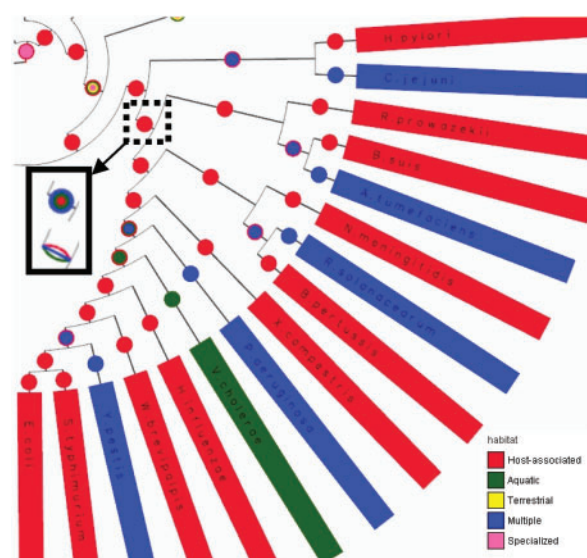
**Fig. 1.** Display of top-down (a), perspective (b) and cone (c) radial trees, showing the mapping of temperature range (green = mesophilic, red = thermophilic, yellow = hyperthermophilic) onto a projected version of the supertree constructed by Beiko *et al.* (2005). The arrow at the base of tree (a) corresponds to the highlighted taxon (*Archaeoglobus fulgidus*) in the menu to the right. Most-parsimonious mappings of character states to internal nodes of the tree are indicated with colored balls: the computed consistency index in this case is 0.286.

ancestral archaeon is suggested to be hyperthermophilic based on a parsimony criterion, whereas hyperthermophilic bacteria appear to be polyphyletic; this distribution may be due to separate origins, LGT or model violation in the process of tree inference. Views may be rotated and translated, and leaves in the tree may be selected in order to display the information for that leaf.

Color-coded character states are mapped to the leaves, and internal edges can be multicolored with ball, arc or split-edge visualizations to show the entire set of descendent traits. There is no limit on the number of colors that can be displayed, other than the obvious upper bound defined by the number of leaves. Taxonomic assignments are presented as arcs on the circumference of the radial tree. In Figure 2, we demonstrate the mapping of species-specific habitat traits from NCBI (<http://www.ncbi.nlm.nih.gov/genomes/lproks.cgi>) to a projected version of the prokaryotic supertree inferred by Beiko *et al.* (2005). Edges may be mapped to show the result of applying Fitch's parsimony algorithm based on a user-selected leaf trait. Supplementary Figures 1–4 show different types of internal edge visualization.

#### 4 CONCLUSIONS

We envision several uses for current and future versions of Radié. One important use will be in the association of different homologous genes and traits (i.e. phylogenetic profiles) with an underlying



**Fig. 2.** Top-down view of the tree relating 14 proteobacteria and the corresponding habitat of each. For internal edges, multicolored balls show the most parsimonious habitat type or set of types associated with that edge, given the states of its children, inferred using the Fitch algorithm. The insets show the complete set (as opposed to the Fitch set) of descendants of the dotted box with a nested ball and multiple arc view.

species phylogeny or hierarchy: quantification of the degree of convexity will suggest vertical or lateral inheritance of certain traits. Displaying trees that optimize the convexity of particular traits would represent hypotheses of the (vertical and lateral) ways in which particular traits might have been inherited.

#### ACKNOWLEDGEMENTS

Kathryn Duffy and Norman MacDonald assisted with layout algorithms and troubleshooting.

*Funding:* Auckland University of Technology DCT/SCMS Research Leave Grant; Genome Atlantic; the Canada Research Chairs program.

*Conflict of Interest:* none declared.

#### REFERENCES

- Beiko, R.G. *et al.* (2005) Highways of gene sharing in prokaryotes. *Proc. Natl Acad. Sci. USA*, **104**, 14332–14337.
- Fitch, W.M. (1971) Toward defining the course of evolution: minimum change for a specified tree topology. *Syst. Zool.*, **20**, 406–416.
- Koonin, E.V. and Wolf, Y.I. (2008) Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world. *Nucleic Acids Res.*, **36**, 6688–6719.
- Letunic, I. and Bork, P. (2007) Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics*, **23**, 127–128.
- Maddison, D.R. *et al.* (1997) NEXUS: an extensible file format for systematic information. *Syst. Biol.*, **46**, 590–621.
- Sapp, J. (2005) The bacterium's place in nature. In Sapp, J. (ed.) *Microbial Phylogeny and Evolution: Concepts and Controversies*. Oxford University Press, New York, pp. 3–52.
- Stechmann, A. and Cavalier-Smith, T. (2002) Rooting the eukaryote tree by using a derived gene fusion. *Science*, **297**, 89–91.