

Ceiba: scalable visualization of phylogenies and 2D/3D image collections

Michael J. Sanderson

Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ 85721, USA

Associate Editor: David Posada

ABSTRACT

Summary: Phylogenetic trees with hundreds of thousands of leaves are now being inferred from sequence data, posing significant challenges for visualization and exploratory analysis. Image data supplying valuable context for species in trees (and cues for exploring them) are becoming increasingly available in biodiversity databases and elsewhere but have rarely been built into tree visualization software in a scalable way. *Ceiba* lets the user explore large trees and inspect image collection arrays (sets of ‘homologous’ images) comprising mixtures of 2D and 3D image objects. *Ceiba* exploits recent improvements in graphics hardware, OpenGL toolkits and many standard high-performance computer graphics strategies, such as texture compression, level of detail control, culling, animations and image caching. Its tree layouts can be tuned by user-provided phylogenetic definitions of subtrees. The code has been extensively tested on phylogenies of up to 55 000 leaves and images.

Availability and implementation: A manual, datasets, source code (distributed under GPL) and binaries for OS X are available at <http://sourceforge.net/projects/ceiba>.

Contact: sanderm@email.arizona.edu

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

Received on December 30, 2013; revised on March 19, 2014; accepted on April 29, 2014

1 INTRODUCTION

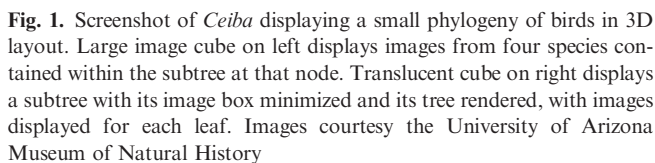
Information visualization can foster exploratory analysis of large and complex datasets. Phylogenetic tree visualization has been an active area of research in the past decade (Page, 2012), with competing programs offering significantly different features, user interfaces and scalability (Block *et al.*, 2012; Hughes *et al.*, 2004; Huson and Scornavacca, 2012; Jordan and Piel, 2008; Kim and Lee, 2007; Munzner *et al.*, 2003; Rosindell and Harmon, 2012; Sanderson, 2006; one website lists 113 programs: http://bioinfo.unice.fr/biodiv/Tree_editors.html; see Supplementary Table S1). In recent years, scalability has assumed greater importance, as phylogenetic trees have markedly increased in size. Trees with ~1000 leaves are now fairly common (e.g. Fabre *et al.*, 2012), and attempts to build trees with 10 000–50 000 or more leaves (Smith *et al.*, 2011) signal the increasing demand for large trees in comprehensive comparative studies in ecology, evolution and other domains (Roquet *et al.*, 2013).

Phylogenetic trees are often annotated with low complexity data, such as discrete traits (e.g. nucleotides), but many biological features, such as morphological traits, are of much higher complexity, and are commonly visualized as 2D images or 3D models reconstructed from CT scans (e.g. McPeck *et al.*, 2008). Although some tree visualization programs allow imbedded images (e.g. Letunic and Bork, 2011; see Supplementary Table S1), image processing capabilities have not been prominently featured in their development, and some basic challenges to their integration into phylogenies remain. For example, in conventional 2D tree layouts, images often either occlude one another or must be greatly reduced in size to avoid this. Scalable integration of image data with tree visualization could foster exploration of comparative patterns in at least two ways: first, by using these data themselves as cues to aid navigation and exploration at large tree scales, and second, to embed these complex data types in an interactive phylogenetic framework that highlights the evolutionary links between the images. The latter may come from large databases of whole organism images, such as Encyclopedia of Life (<http://eol.org>), databases of specific traits, such as MorphBank (<http://www.morphbank.net>), of 3D models derived from CT scans, such as DigiMorph (<http://www.digimorph.org>).

2 METHODS

Ceiba originated in the *Paloverde* tree visualization code (Sanderson, 2006), but *Ceiba*’s new layout schema combines a much more scalable hierarchical decomposition of large trees with code re-engineered to integrate image data. *Ceiba* ‘layouts’ are of two kinds. The first specifies how the input tree is decomposed into a hierarchy of subtrees, determined automatically or by user-supplied phylogenetic definitions of subtrees. The second concerns the *visual* appearance of these subtrees, how they relate to each other when rendered and their integration with image data. Subtrees can be rendered either as conventional 2D trees like layouts available in other software, or as 3D objects that look like botanical ‘trees’. Subtrees are combined in two visual layout schemes. In the ‘3D’ scheme, *all* subtrees are displayed as 3D objects. Navigation is effected by selecting subtrees and recursively zooming into this defined 3D space. In the ‘fan’ scheme, the root subtree is rendered as a 3D tree, but all descendant subtrees are rendered as 2D semicircular trees (keeping with a botanical metaphor; these are rendered to look something like the planar leaves of a fan palm). Navigation to subtrees in this scheme uses a different metaphor: selecting subtrees maximizes a subtree’s fan, displaying it in the same plane as its parent subtree fan, allowing a user to see several hierarchical levels at once.

A collection of 2D digital images and/or 3D surface models can be associated with any leaf in the tree. Arranged in a ‘homology array’, images at the same spatial position in different leaves can thus represent



Although a few programs are scalable to the same degree and can incorporate images (e.g. Archaeopteryx: <http://www.phylosoft.org/archaeopteryx/>; Block *et al.*, 2012 and Supplementary Table S1), *Ceiba* uses a number of techniques to integrate this scalability with imbedded images. Many of these take advantage of the hierarchical decomposition of the tree to avoid expensive layout computation and rendering operations (of both trees and images) until they are needed. Culling is used extensively, and an optional image caching strategy is implemented in which recently visited but no longer visible images are freed from memory.

Scalability of *Ceiba* was evaluated in depth on two large trees included in the software distribution (839 and 55473 leaves). To evaluate what a user would experience in a lengthy interactive session, the program was run in ‘tour’ mode, an animation that

Block, F. *et al.* (2012) The DeepTree exhibit: visualizing the tree of life to facilitate informal learning. *IEEE Trans. Vis. Comput. Graph.*, **18**, 2789–2798.

Fabre, P.-H. *et al.* (2012) A glimpse on the pattern of rodent diversification: a phylogenetic approach. *BMC Evol. Biol.*, **12**, 88.

Hughes, T. *et al.* (2004) Visualizing very large phylogenetic trees in three dimensional hyperbolic space. *BMC Bioinformatics*, **5**, 1–6.

Huson, D.H. and Scornavacca, C. (2012) Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. *Syst. Biol.*, **61**, 1061–1067.

Jordan, G.E. and Piel, W.H. (2008) PhyloWidget: web-based visualizations for the tree of life. *Bioinformatics*, **24**, 1641–1642.

Kim, N. and Lee, C. (2007) Three-dimensional phylogeny explorer: distinguishing paralogs, lateral transfer, and violation of “molecular clock” assumption with 3D visualization. *BMC Bioinformatics*, **8**, 213.

Letunic, I. and Bork, P. (2011) Interactive tree of life v2: online annotation and display of phylogenetic trees made easy. *Nucleic Acids Res.*, **39**, W475–W478.

McPeck, M.A. *et al.* (2008) The tempo and mode of three-dimensional morphological evolution in male reproductive structures. *Am. Nat.*, **171**, E158–E178.

Munzner, T. *et al.* (2003) TreeJuxtaposer: scalable tree comparison using focus+context with guaranteed visibility. *ACM Trans. Graph.*, **22**, 453–462.

Page, R.D.M. (2012) Space, time, form: viewing the tree of life. *Trends Ecol. Evol.*, **27**, 113–120.

Roquet, C. *et al.* (2013) Building megaphylogenies for macroecology: taking up the challenge. *Ecography*, **36**, 13–26.

Rosindell, J. and Harmon, L.J. (2012) OneZoom: a fractal explorer for the tree of life. *PLoS Biol.*, **10**, e1001406.

Sanderson, M.J. (2006) *Paloverde*: an OpenGL 3D phylogeny browser. *Bioinformatics*, **22**, 1004–1006.

Smith, S.A. *et al.* (2011) Understanding angiosperm diversification using small and large phylogenetic trees. *Am. J. Bot.*, **98**, 404–414.