*Bioinformatics*, YYYY, 0–0

doi: 10.1093/bioinformatics/xxxxx

Advance Access Publication Date: DD Month YYYY

Applications Note

|  |
| --- |
| Phylogenetics  PhyloBrowse: A visual tool to explore phylogenomic data  Corresponding Author1,\*, Co-author2 and Co-Author2  1Department of XXXXXXX, Address XXXX etc., 2Department of XXXXXXX, Address XXXX etc.  \*To whom correspondence should be addressed.  Associate Editor: XXXXXXX  Received on XXXXX; revised on XXXXX; accepted on XXXXX  Abstract  **Summary:** PhyloBrowse is an interactive web-based tool that supports phylogenomic data exploration. Given a phylogenetic tree, it depicts the relationship between individual genes in the supermatrix used for phylogeny reconstruction and nodes within the tree. The data layers can be filtered in a variety of ways to highlight phylogenetic features of interest. *Gil, please add: The link … contains an example Phylobrowse instance for plants.*  **Availability and implementation:** PhyloBrowse is a JavaScript program. Source code, documentation and example files are freely available under the MIT License at https://github.com/???.  **Contact:** example@example.org (Which one??)  **Supplementary information:** Supplementary data are available at *Bioinformatics* online. |

# Introduction

Publically available genomic and transcriptomic data for myriad species and efficient computational methods allow the robust evolutionary reconstruction of various parts of the tree of life are becoming possible (Lee et al., 2011; Zapata et al., 2015, Silva et al., 2015 and Shen et al., 2016). ~~Phylogenomic analysis can be used to identify orthologous genes among sets of taxa and to predict the functions of uncharacterized genes based on their phylogenetic relatedness to well-characterized genes (Eisen, 1998).~~

~~Of critical importance, phylogenomics lays the foundation for downstream functional and molecular evolutionary analyses that focus on identifying genes underlying major phenotypic traits (Clark et al., 2007; Lee et al., 2011; Silva et al., 2015 and Pease et al., 2016). ).~~ Phylogenomic analyses can estimate a quantitative measurement relating genes to specific branches or nodes within a phylogenetic tree, such as molecular selection (non-synonymous to synonymous rate ratio, best known as dN/dS) and the Partitioned Branch Support (PBS, also referred to as Partitioned Bremer Support). Numerous tools have been developed for phylogenetic tree visualization and annotation, such as PhyD3 (Kreft et al., 2017), ETE Toolkit (Huerta-Cepas et al., 2010; Huerta-Cepas et al., 2016), EvolView (He et al., 2016), iTOL (Letunic and Bork, 2016), Dendroscope (Huson and Scornavacca, 2012) and FigTree (<http://tree.bio.ed.ac.uk/software/figtree/)>. While effective at displaying a tree topology, and static node and branch labels, current tools lack an interactive, exploration and mining interface for genome-wide gene-to-node relationship data.

# Description and Features

PhyloBrowse is an interactive web-based data visualization tool for exploring genome-scale phylogenetic trees and their associated gene-to-node measurements. PhyloBrowse allows dynamic visualization of sets of genes (i.e., processes) that support clade separation at one or more internal nodes. We demonstrate the salient features of PhyloBrowse using a previously-published genome-scale phylogenomic analysis by Lee et al., 2011 (Fig. 1). In this example dataset, PBS support values were calculated for every gene at every internal node in the tree. GO term enrichment analysis was also reported per node, indicating enriched terms for genes that provide positive support (PBS ≥4) to a given node. Although we refer here to PBS values and GO term annotations, many gene-to-node associations and annotation types can be explored using PhyloBrowse (e.g., KEGG pathways, InterPro domains, etc.).

## Tree visualization

PhyloBrowse displays a phylogenetic tree provided in a Newick format as either a cladogram or a phylogram (if branch lengths are provided) ~~using a toggle button~~. Nodes are sized by their average PBS value. Each node is associated with a tooltip window that shows the list of genes and their PBS values in brackets, as well as any enriched GO terms.

## Node filters

Filters can be applied to direct the research at nodes of interest. For example, using the “Gene” or the “GO Term” filters, the user can identify nodes containing genes and/or GO terms of interest, respectively. The selected genes or GO terms are listed in the “Filter Information” panel. Nodes that contain any of the selected genes or GO terms (associated with one or more genes) in their list, will be re-colored to orange, while other nodes will remain green or collapse if none of the child nodes contain the genes/terms. If selected GO terms are enriched in a particular node, that node will be re-colored in purple. The nodes can also be filtered by a range of the PBS values, or manually (using a shift+click). The list of GO terms can be filtered prior to their selection using the “Limit GO terms by p-value” slider.

Fig. 1. The PhyloBrowse interface for exploring genome-scale phylogenetic trees and their associated gene-to-node measurements. For demonstration purposes, data was taken from a seed plants phylogenomic analysis, by Lee et al., 2011. This PhyloBrowse session is available at <http://phylobrowse.bio.nyu.edu/>. Tree visualization – Each node in the tree is associated with a tooltip window containing the genes (and gene-to-node values) and the enriched annotation terms (and p-values) that are associated with them. Node filters panel – Genes and annotation terms can be selected to highlight (in orange) nodes that are associated with them. Nodes enriched for selected terms are colored in purple. Other nodes will remain green or collapse (in black). Union/Intersect/Difference – Selected nodes are colored in light blue, as the lists of genes/terms are dynamically updated for each category. Taxonomy classification – a dropdown menu in the top right corner allows to switch between taxonomic ranks that are further displayed adjacent to the tree terminals. Every taxonomic group is assigned with a different color, which allows detecting incongruence placements along the tree.

## GO terms/node heatmap

When the “GO Term Filter” is applied, PhyloBrowse generates a heatmap showing all enriched GO terms for highlighted nodes (colored in either orange or purple). ~~(Darker colors represent smaller p-values).~~

## Union/Intersect/Difference of genes and GO term lists

PhyloBrowse allows users to compare genes or GO term lists between any two or more internal nodes, which can be useful if looking for genes that may have converged due to shared life histories (e.g., drought tolerance). Selected nodes are colored in light blue.

## Taxonomy classification

The taxonomy classification feature allows selection of the taxonomic ranks displayed adjacent to the tree terminals. Every unique taxonomic name is assigned a different color, providing a quick identification of any incongruence between tree topology and the taxonomic classification.

# Implementation

PhyloBrowse is written in JavaScript and may be used on an http server or on a local host. PhyloBrowse uses the D3 JavaScript library (https://d3js.org/), and the D3 Phylogram plug-in (https://gist.github.com/1036776) for layout and styling of the graph. It uses the jQuery library for element manipulation, user interaction and several elements such as sliders. It uses the JavaScript Newick format parser (https://github.com/jasondavies/newick.js) to parse the input tree. The inputs for PhyloBrowse are indicated in a config file and are in JSON format (example files can be found in Supplemental S1):

~~A node\_info.json file (required) that indicates node support values (e.g. Bootstrap values), the node gene lists and the associated gene-to-node values, as well as any enriched annotations (e.g., GO terms) that are associated with the node gene list.~~

~~A taxonomy.json file (recommended), with the defined taxonomy levels and names for each taxon.~~

~~A gene identifier to annotation identifier mapping file (gene2Annot.json) (recommended), and an annotation identifier to annotation description mapping file (AnnotDescription.json). These files should include only annotations that are associated with the input gene identifiers, to reduce loading speed.~~

~~Instructions as well as a set of python scripts to generate the above input files are provided with the software.~~

# Conclusions

PhyloBrowse offers researchers an intuitive, fast and an interactive way to explore their phylogenomic results. PhyloBrowse sessions can be easily shared with colleagues, or be publically made available for the scientific community. Using the various filters and the node color codes, researchers can immediately identify where a specific node or function have become evolutionary divergent. Researchers can then study the gene-to-node values and retrieve genes co-occurring with a trait of interest. PhyloBrowse is open-source.

Funding

This work has been supported by the National Science Foundation [IOS-0922738 to G.C.].

*Conflict of Interest:* none declared.

References

Clark,A.G. et al. (2007) Evolution of genes and genomes on the Drosophila phylogeny. *Nature*, 450, 203-218.

Eisen,J.A (1998) Phylogenomics: Improving functional predictions for uncharacterized genes by evolutionary analysis. *Genome Res.*, 8, 163–167.

He,Z. et al. (2016) EvolView v2: an online visualization and management tool for customized and annotated phylogenetic trees. *Nucleic Acids Res.*, 44, W236–W241.

Huerta-Cepas,J. et al. (2016) ETE 3: reconstruction, analysis, and visualization

of phylogenomic data. *Mol. Biol. Evol.*, 33, 1635–1638.

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

Kreft,L. et al. (2017) PhyD3: a phylogenetic tree viewer with extended phyloXML support for functional genomics data visualization. *Bioinformatics*, 33, 2946–2947.

Lee,E. et al. (2011) A Functional Phylogenomic View of the Seed Plants. PLoS Genet., 12, e1002411.

Letunic,I. and Bork,P. (2016) Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.*, 44, W242-W245.

Pease,J.B. et al. (2016) Phylogenomics Reveals Three Sources of Adaptive Variation during a Rapid Radiation. *PLoS Biol.*, 14, e1002379.

Shen,X.X. et al. (2016) Reconstructing the Backbone of the Saccharomycotina Yeast Phylogeny Using Genome-Scale Data. *G3 (Bethesda)*, 6, 3927–3939.

Silva,D.N. et al. (2015) Genomic Patterns of Positive Selection at the Origin of Rust Fungi. *PLoS ONE*, 10, e0143959.

Zapata,F. et al. (2015) Phylogenomic Analyses Support Traditional Relationships within Cnidaria. *PLoS ONE*, 10, e0139068.