

PhyloJIVE: Integrating biodiversity data with the Tree of Life

Garry Jolley-Rogers¹, Temi Varghese¹, Paul Harvey¹, Nick dos Remedios² and Joseph T. Miller^{1,*}

¹Centre for Australian National Biodiversity Research, CSIRO Plant Industry, Canberra ACT 2601, Australia and ²Atlas of Living Australia, CSIRO Ecosystem Sciences, GPO Box 1700, Canberra, ACT 2601, Australia

Associate Editor: Martin Bishop

ABSTRACT

Motivation: Rich collections of biodiversity information such as spatial distributions, species descriptions and trait data are now synthesized in publicly available online sources such as GBIF. Also phylogenetic knowledge now provides a sound understanding of the origin of organisms and their place in the tree of life. We demonstrate with PhyloJIVE that any phylogenetic tree can be linked to online biodiversity data in the browser. This evolutionary view of biodiversity data is demonstrated in a case study that suggests that this approach may be useful to scientists and non-experts users.

Contact: joe@acaciamulga.net

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

Received on September 3, 2013; revised on December 23, 2013; accepted on January 12, 2014

1 INTRODUCTION

The growing catalogue of online biodiversity data sources has reached a scale where integration with phylogenies can inform comparative biology research in morphology, ecology, biogeography, disease resistance, ecosystem services, food security and climate change (Chapman, 2005; Cracraft, 2002; Miller and Jolley-Rogers, 2014). There are now mature services (e.g. GBIF, EOL and the ALA) that provide biodiversity data, including species profiles, morphological, distribution, images and taxonomic data, to researchers and the general public (Hardisty and Roberts, 2013). These data sources provide data in taxon profiles but not through phylogenies.

The research community has generated DNA-based phylogenetic trees (Pace *et al.*, 2012), and many are stored in repositories such as TreeBase II (<http://www.treebase.org/>). Many data analysis and visualization suites are now available, which produce and display phylogenies (Pavlopoulos *et al.*, 2010) but none integrate data, such as morphological character states derived from web services, on a phylogeny. Here we demonstrate PhyloJIVE (Phylogeny Javascript Information Visualiser and Explorer), a web-based application that integrates biodiversity information aggregated from many sources with phylogenetic trees. PhyloJIVE complements other initiatives that provide visualizations such as Phylowidget, OneZoom (<http://www.onezoom.org>) and jsphylosvg (Jordan and Piel, 2008; Page, 2011; Pavlopoulos *et al.*, 2010; Smits and Ouerney, 2010). We do not seek to

replicate what they do well; rather, we provide a convenient means to assemble online data in a phylogenetic context.

2 METHODS

PhyloJIVE is an open source web application consisting of code and templates designed so that any website can provide a phylogenetic view of biodiversity data. Source code is available via github (<http://trin.github.io/phyloJIVE/>), and the ALA (<http://phylojive.ala.org.au>) hosts an example implementation. Designed to be flexible, it operates as a client-side web application that renders phylogenies on a HTML5 canvas. Matters such as user authentication and access control are left to the host website. Written in JavaScript, it requires no plug-ins. Performance is determined by the viewer's computer. PhyloJIVE has three components: *PhylogenyExplorer*, *PhyloJIVE-templates* and *charJSON*. Phylogeny Explorer is an extension to the JIT (Javascript Infovis Toolkit library) that adds a novel, compact tree visualization with variable length branches. Actions and online data sources can be linked to every node.

PhyloJIVE templates demonstrate the use of the visualization and provide standardized tests to verify algorithms and code. It demonstrates how to navigate the tree and component clades; to adjust the visualization and other parameters; to select taxon character states; to upload of trees in newick form and character sets in CSV or charJSON form; to fetch character sets from Identify Life; and to display the results of web service queries (e.g. species pages, maps, images) on the tree. Sets of up to 10 quantitative or qualitative characters can be mapped onto the tree. Character states are reconstructed using simple algorithms (see Supplemental information), a summary of which appears in the application when the mouse pointer hovers over a node.

charJSON is a data structure designed for biodiversity web services that delivers qualitative and quantitative traits/characters. It can be parsed by most programming languages. Supplementary Table S1 shows a formal description of charJSON in Extended Backus-Naur Form.

3 CASE STUDY

An implementation of PhyloJIVE (<http://phylojive.ala.org.au/treeViewer/show/Case+Study+1%3A+Acacia>) was developed on the Atlas of Living Australia. It was used to integrate spatial data points from the ALA with user-uploaded data: an *Acacia* phylogeny (Miller *et al.*, 2011) and traits in a.csv file. Uniquely for *Acacia*, PhyloJIVE also integrated data from the IdentifyLife (<http://www2.identifylife.org>) web services plant identification key, Wattle (Maslin, 2001), which contains a morphological character matrix of *Acacia* species. After uploading the data, we selected 10 characters and mapped character states onto the tree, while also automatically reconstructing ancestral states for each node. Branches were coloured by the character states of the first character selected (Fig. 1A).

*To whom correspondence should be addressed.

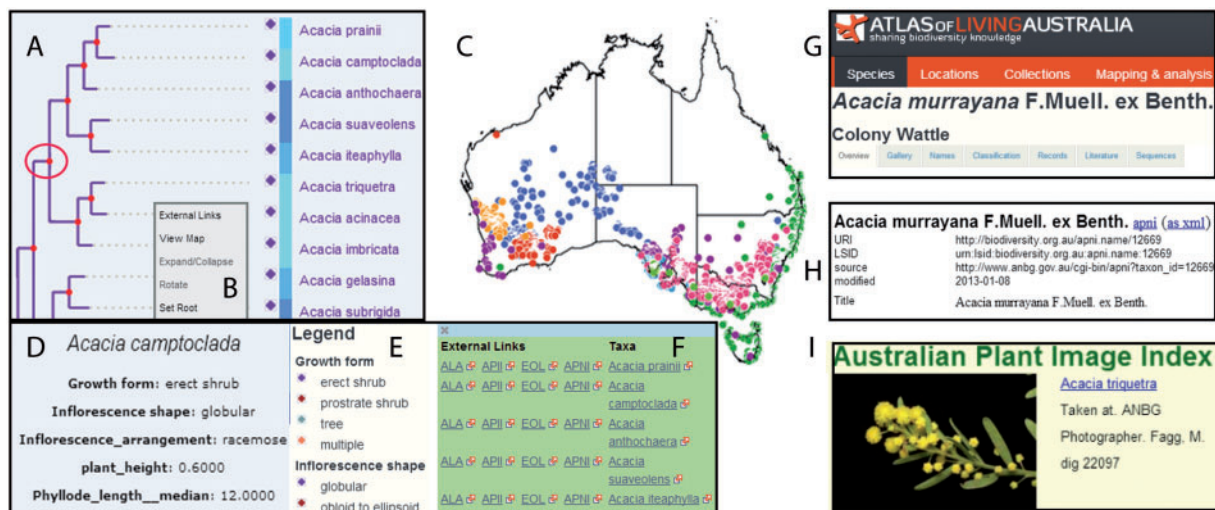


Fig. 1. (A) Phylogeny; (B) menu that appears as the pointer hovers; (C) spatial distribution map for clade; (D) character states and (E) legend; (F) linked data sources; (G) Atlas of Living Australia, (H) Australian Plant Image Index, (I) National Species List

Users are able to interact with the tree and the biodiversity data through the web application. They can view multiple phylogenetic depths and collapse their view of select clades (Fig. 1A). They can view allied data for terminals or nodes by left-clicking in the application (Fig. 1D-I). Assembled web services links included nomenclature, species pages and maps from the ALA. Users can view distribution maps for a single terminal or clades of up to 32 taxa, with each terminal coloured differently (Fig. 1C). Maps can also show the distribution of any available character for a clade (Supplementary Fig. SI C-F). Details of reconstructed states for all selected characters appear as pointer hovers over nodes and terminals (Fig. 1E). A matrix alongside the tree summarized character states with symbols for qualitative and a heat map for quantitative characters (Fig. 1A). Users are able to select a clade and export the localities to the geospatial portal of the ALA (see Supplementary Fig. SI) to investigate spatial correlations with environmental layers.

4 CONCLUSIONS

The ALA implementation is set to access primarily Australian data sources. The means to discover and use web services is evolving and applied inconsistently in the community of biodiversity data providers (Pedinaci and Domingue, 2010; Parr *et al.*, 2011). Other data sources can be added as required; however, further development is required for data sources to be added easily in the form of user-friendly pull-down menus.

The simple visualization of multiple character states on a tree is novel to PhyloJIVE, and users can identify interesting trends that can form hypotheses that require further investigation. Because PhyloJIVE is a web-based tool, it is easily shared among workers. A worker who wants to share an interesting discovery can send the URL to a colleague who can open PhyloJIVE, reload the datasets and call up the URL. Likewise, a user can bookmark a particular character setup for later investigation.

PhyloJIVE demonstrates that it is possible and useful to assemble interactive biodiversity data from diverse sources. Viewed through phylogeny, the evolutionary context of morphological,

spatial and ecological data becomes clear and accessible to non-specialists.

ACKNOWLEDGEMENTS

The authors acknowledge the Taxonomy Research & Information Network (TRIN), the Centre for Australian National Biodiversity Research (CANBR) and the Atlas of Living Australia (ALA).

Conflict of Interest: none declared.

REFERENCES

- Chapman, A.D. (2005) *Uses of Primary Species Occurrence Data Global Biodiversity Information Facility*. Copenhagen, Denmark, pp. 1–111.
- Cracraft, J. (2002) The seven great questions of systematic biology: an essential foundation for conservation and the sustainable use of biodiversity. *Ann. Missouri Bot. Gard.*, **89**, 127–144.
- Hardisty, A. and Roberts, D. (2013) A decadal view of biodiversity informatics: challenges and priorities. *BMC Ecol.*, **13**, 16.
- Jordan, G.E. and Piel, W.H. (2008) PhyloWidget: web-based visualizations for the tree of life. *Bioinformatics*, **24**, 1641–1642.
- Maslin, B.R. (2001) Wattle, Acacias of Australia. CSIRO PUBLISHING/ Australian Biological Resources Study (ABRS).
- Miller, J.T. *et al.* (2011) The evolution and phylogenetic placement of invasive *Acacia* species. *Divers. Distrib.*, **17**, 848–860.
- Miller, J.T. and Jolley-Rogers, G. (2014) Correcting the disconnect between phylogenetics and biodiversity informatics. *Zootaxa*, **3754**, 195–200.
- Pace, N.R. *et al.* (2012) Classic perspective: phylogeny and beyond: scientific, historical, and conceptual significance of the first tree of life. *Proc. Natl Acad. Sci. USA*, **109**, 1011–1018.
- Page, R.D.M. (2011) Space, time, form: viewing the Tree of Life. *Trends Ecol. Evol.*, **27**, 113–120.
- Parr, C.S. *et al.* (2011) Evolutionary informatics: unifying knowledge about the diversity of life. *Trends Ecol. Evol.*, **27**, 1–10.
- Pavlopoulos, G.A. *et al.* (2010) A reference guide for tree analysis and visualization. *BioData Min.*, **3**, 1.
- Pedinaci, C. and Domingue, J. (2010) Toward the next wave of services: linked services for the web of data. *J. Univers. Comput. Sci.*, **16**, 1694–1719.
- Smits, S.A. and Ouverney, C.C. (2010) jsPhyloSVG: a javascript library for visualizing interactive and vector-based phylogenetic trees on the web. *PLoS One*, **5**, e12267.