

RESEARCH NOTE

Open Access



Treehouse: a user-friendly application to obtain subtrees from large phylogenies

Jacob L. Steenwyk and Antonis Rokas*

Abstract

Objective: Phylogenetic trees that contain hundreds to thousands of taxa are now routinely generated. Retrieving the relationships among a subset of taxa in these large phylogenies can be a challenging or time-consuming task. Addressing this challenge requires the development of tools that facilitate the easy retrieval of subtrees from any user-specified set of taxa in a given phylogeny.

Results: We developed *treehouse*, an open source tool that enables the retrieval of any subtree from a given large phylogeny. With a three-step workflow, *treehouse* successfully allows a user to obtain a subtree from any phylogeny. *Treehouse* can help researchers to explore the relationships among any set of taxa from across the tree of life. *Treehouse* is implemented as a shiny application in the R programming language. *Treehouse* software and usage instructions are publicly available at <https://github.com/JLSteenwyk/treehouse>.

Keywords: Phylogenomics, Phylogenetics, Big data, Tree, Tree pruning, Shiny, Graphical user interface

Introduction

Evolutionary biology relies on understanding the phylogenetic relationships among sets of genes, traits, and organisms under investigation. However, large phylogenies that contain hundreds of taxa are increasingly becoming inaccessible to researchers interested in the relationships of just a few representatives. For example, some phylogenies are so large that taxon information is often challenging or impossible to visualize and is often excluded [1–4]; similarly, the lengths of many internal branches are often very short and the constraints of displaying a large tree in a letter-sized page make the tracing of relationships among a subset of taxa challenging and unnecessarily time-consuming. These issues will increase in frequency as the numbers of taxa included in phylogenies of genes, metagenomes, genomes, etc. continues to rapidly rise.

To address these issues, we introduce *treehouse*, a user-friendly application with minimal dependencies that facilitates the retrieval of subtrees from any

user-specified set of taxa in a given phylogeny. Our simple three-step workflow allows users to obtain subtrees from a curated and growing database of large-scale phylogenetic trees from across the tree of life. Additionally, users may obtain subtrees from their own phylogenies which, can facilitate data exploration and inter-disciplinary collaboration. For easy integration into pre-existing project workflows, subtrees obtained from *treehouse* can be easily be downloaded as a newick file or PDF file that retains branch length information. *Treehouse* enables beginner and expert evolutionary biologists alike to reap the benefits of large-scale phylogenetic projects and use them to test evolutionary-based hypotheses.

Main text

Materials and methods

Data acquisition

The *treehouse* contains a database of 20 representative large phylogenies from across the tree of life (Table 1).

Description of the software

Using *treehouse* requires the R packages PHYTOOLS, version 0.6–60 [21], and SHINY, version 1.2.0 (<https://shiny.rstudio.com/>). Dependencies of PHYTOOLS

*Correspondence: antonis.rokas@vanderbilt.edu
Department of Biological Sciences, Vanderbilt University, Nashville, TN 37235, USA



Table 1 Curated phylogenies currently available in *treehouse's* database

Highest level of taxonomic organization	Taxon or taxa represented	Number of taxa	References
Animals	Birds	198 taxa	[5]
Animals	Birds	48 taxa	[6]
Animals	Insects	144 taxa	[7]
Animals	Mammals	37 taxa	[8]
Animals	Mammals	36 taxa	[9]
Animals	Metazoans	36 taxa	[10]
Animals	Metazoans	70 taxa	[11]
Animals	Vertebrates	58 taxa	[12]
Animals	Worms	100 taxa	[13]
Fungi	<i>Aspergillus</i> and <i>Penicillium</i>	81 taxa	[14]
Fungi	<i>Cryptococcus neoformans</i>	387 strains	[15]
Fungi	Fungi	214 taxa	[16]
Fungi	Agaricomycetes	5284 taxa	[2]
Fungi	<i>Saccharomyces cerevisiae</i>	1011 strains	[1]
Fungi	Saccharomycotina	86 taxa	[17]
Fungi	Saccharomycotina	332 taxa	[4]
Plant	Caryophyllales	95 taxa	[18]
Plant	Flowering plants	45 taxa	[19]
Plant	Land plants	103 taxa	[20]
Tree of life	Tree of life	3083 taxa	[3]

includes MAPS, version 3.3.0 (<https://cran.r-project.org/web/packages/maps/index.html>), and APE, version 5.3 [22]. To present the phylogeny as depicted by the original authors, phylogenies from *treehouse's* database are rooted. The taxa chosen to root the phylogeny on are inferred from figures presented in the original manuscript or, in the case of phylogenies presented without taxa names, personal communications with the authors. Phylogenies are rooted using PHYTOOLS's root() function. Using the list of taxa provided by the user, *treehouse* determines the list of taxa to remove from the phylogeny using the setdiff() function. The resulting list is then used to remove taxa in the phylogeny using PHYTOOLS's drop.tip() function. To write out the resulting phylogeny in a newick-formatted text file or display it in a scalable-vector-graphic-formatted pdf file, we use the write.tree() and plot.phylo() functions in APE, respectively. To create a user-friendly and intuitive user-interface, we used SHINY.

Results

A three-step workflow to obtain subtrees

Treehouse is designed to have a simple user-interface that guides a user through an intuitive three-step workflow (Fig. 1A) and user interface (Fig. 1B).

1. Tree selection

A user can choose between five tabs—userTree, Animals, Fungi, Plants, and Tree of Life—located at the top of the user interface (Fig. 1Ba). When using phylogenies from the *treehouse* database, a user selects the desired phylogeny using a dropdown menu (Fig. 1Bi; left). In userTree, a user selects a phylogeny in newick format from their local computer (Fig. 1Bi; right).

2. Selection of Taxa

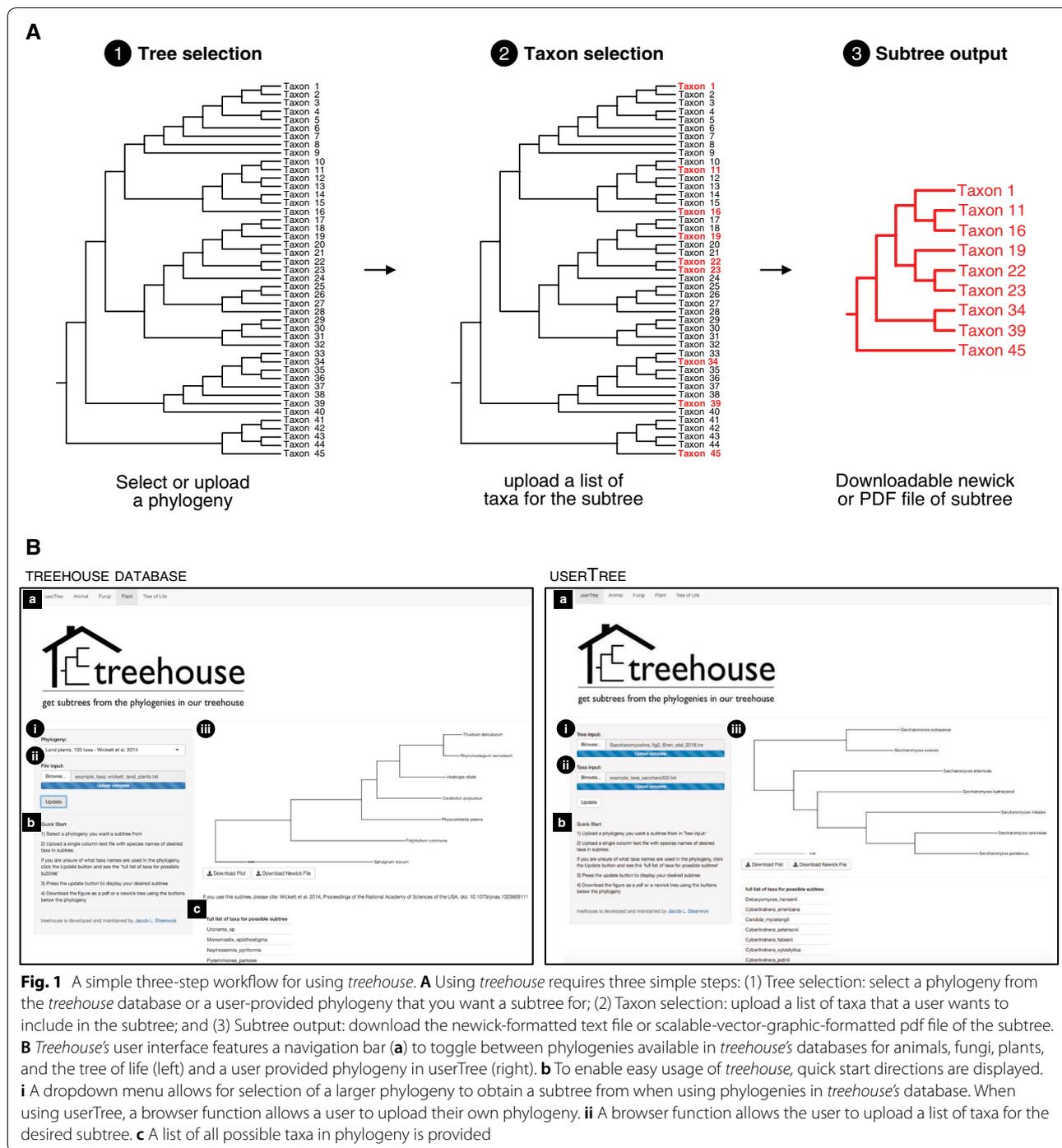
A user next uploads a text file containing the single-column list of taxa that they want a subtree for (Fig. 1Bii). Here, each taxon name must be identical to a taxon name in the full phylogeny.

3. Subtree output

By clicking the 'Update' button, the user launches *treehouse* subtree retrieval. The subtree is plotted to the right of the side panel and buttons that allow the user to download a pdf or text file of the subtree are below it (Fig. 1Bii). Lastly, the full set of taxa in the currently uploaded *treehouse* phylogeny is listed (Fig. 1Bc; left).

Conclusion

Treehouse is a simple and powerful tool that facilitates subtree retrieval from large phylogenies.



Limitations

Treehouse's functionality rests on the performance of one task, namely removing taxa from a phylogeny. To the experienced phylogenetic or phylogenomic researcher, this might seem to be a trivial task but is not so for most users of phylogenetic trees and no other user-friendly

methods are available. Thus, we anticipate the 'typical' *treehouse* users to be researchers that use phylogenies to form hypotheses but do not routinely infer phylogenies themselves. We also anticipate *treehouse* to be a useful teaching tool.

Acknowledgements

We thank members of the Rokas laboratory, and in particular Xing-Xing Shen, for helpful comments, discussion, and user feedback. We thank Christina Cuomo for providing useful user feedback. We thank the research community for their input on what phylogenies they would like to be part of the *treehouse* database.

Authors' contributions

JLS and AR conceived the research and wrote the article. JLS conducted the research and implemented *treehouse* in the R programming language. Both authors read and approved the final manuscript.

Funding

JLS was supported by the Graduate Program in Biological Sciences at Vanderbilt University. AR was supported by the National Science Foundation (DEB-1442113), the Burroughs Wellcome Fund, and the Guggenheim Foundation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Availability of data and materials

All data, materials, and code are publicly available at <https://github.com/JLSteenwyk/treehouse>.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 23 May 2019 Accepted: 21 August 2019

Published online: 27 August 2019

References

- Peter J, De Chiara M, Friedrich A, Yue J-X, Pflieger D, Bergström A, et al. Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates. *Nature*. 2018;556:339–44. <https://doi.org/10.1038/s41586-018-0030-5>.
- Varga T, Krizsán K, Földi C, Dima B, Sánchez-García M, Sánchez-Ramírez S, et al. Megaphylogeny resolves global patterns of mushroom evolution. *Nat Ecol Evol*. 2019;3:668.
- Hug LA, Baker BJ, Anantharaman K, Brown CT, Probst AJ, Castelle CJ, et al. A new view of the tree of life. *Nat Microbiol*. 2016;1:16048. <https://doi.org/10.1038/nmicrobiol.2016.48>.
- Shen X-X, Opulente DA, Kominek J, Zhou X, Steenwyk JL, Buh KV, et al. Tempo and mode of genome evolution in the budding yeast subphylum. *Cell*. 2018;175(1533–1545):e20.
- Prum RO, Berv JS, Dornburg A, Field DJ, Townsend JP, Lemmon EM, et al. A comprehensive phylogeny of birds (*Aves*) using targeted next-generation DNA sequencing. *Nature*. 2015;526:569–73.
- Jarvis ED, Mirarab S, Aberer AJ, Li B, Houde P, Li C, et al. Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science*. 2014;346:1320–31.
- Misof B, Liu S, Meusemann K, Peters RS, Donath A, Mayer C, et al. Phylogenomics resolves the timing and pattern of insect evolution. *Science*. 2014;346:763–7.
- Song S, Liu L, Edwards SV, Wu S. Resolving conflict in eutherian mammal phylogeny using phylogenomics and the multispecies coalescent model. *Proc Natl Acad Sci*. 2012;109:14942–7.
- Tarver JE, dos Reis M, Mirarab S, Moran RJ, Parker S, O'Reilly JE, et al. The interrelationships of placental mammals and the limits of phylogenetic inference. *Genome Biol Evol*. 2016;8:330–44.
- Borowiec ML, Lee EK, Chiu JC, Plachetzki DC. Extracting phylogenetic signal and accounting for bias in whole-genome data sets supports the Ctenophora as sister to remaining Metazoa. *BMC Genomics*. 2015;16:987.
- Whelan NV, Kocot KM, Moroz LL, Halanych KM. Error, signal, and the placement of Ctenophora sister to all other animals. *Proc Natl Acad Sci*. 2015;112:5773–8.
- Chen M-Y, Liang D, Zhang P. Selecting question-specific genes to reduce incongruence in phylogenomics: a case study of jawed vertebrate backbone phylogeny. *Syst Biol*. 2015;64:1104–20.
- Struck TH, Golombek A, Weigert A, Franke FA, Westheide W, Purschke G, et al. The evolution of annelids reveals two adaptive routes to the interstitial realm. *Curr Biol*. 2015;25:1993–9.
- Steenwyk JL, Shen X-X, Lind AL, Goldman GH, Rokas A. A robust phylogenomic time tree for biotechnologically and medically important fungi in the genera *Aspergillus* and *Penicillium*. *MBio*. 2019;10:e00925.
- Desjardins CA, Giamberardino C, Sykes SM, Yu C-H, Tenor JL, Chen Y, et al. Population genomics and the evolution of virulence in the fungal pathogen *Cryptococcus neoformans*. *Genome Res*. 2017;27:1207–19.
- James TY, Kauff F, Schoch CL, Matheny PB, Hofstetter V, Cox CJ, et al. Reconstructing the early evolution of Fungi using a six-gene phylogeny. *Nature*. 2006;443:818–22.
- Shen XX, Zhou X, Kominek J, Kurtzman CP, Hittinger CT, Rokas A. reconstructing the backbone of the Saccharomycotina yeast phylogeny using genome-scale data. *G3 Genes[Genomes]Genetics*. 2016;6:3927–39.
- Yang Y, Moore MJ, Brockington SF, Soltis DE, Wong GKS, Carpenter EJ, et al. Dissecting molecular evolution in the highly diverse plant clade Caryophyllales using transcriptome sequencing. *Mol Biol Evol*. 2015;32:2001–14.
- Xi Z, Liu L, Rest JS, Davis CC. Coalescent versus concatenation methods and the placement of Amborella as sister to Water Lilies. *Syst Biol*. 2014;63:919–32.
- Wickett NJ, Mirarab S, Nguyen N, Warnow T, Carpenter E, Matasci N, et al. Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proc Natl Acad Sci*. 2014;111:E4859–68.
- Revell LJ. phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol Evol*. 2012;3:217–23.
- Paradis E, Claude J, Strimmer K. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*. 2004;20:289–90.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

