**TinselR – An R Shiny Application for Annotating Phylogenetic Trees**

Running Title: TinselR for annotating trees

Jennafer A. P. Hamlincurrent,1,2,\*, Teofil Nakov3, and Amanda Williams Newkirk2

**1** Association of Public Health Laboratories Bioinformatics, Silver Springs, MD, USA

**2** Enteric Diseases Laboratory Branch, Centers for Disease Control and Prevention, Atlanta, GA, USA

**3** Department of Biological Sciences, University of Arkansas, Fayetteville, Arkansas, USA

**\*** Address correspondence to Jennafer A. P. Hamlin: [ptx4@cdc.gov](mailto:ptx4@cdc.gov)

**current** Respiratory Diseases Laboratory Branch, Centers for Disease Control and Prevention, Atlanta, GA, USA

**ABSTRACT**

Public health laboratories obtain whole-genome sequences of pathogens to confirm outbreaks and identify transmission routes. Here we present **tinselR,** an open-source and user-friendly application for visualization and annotation of relatedness among pathogens with phylogenetic trees. By combining annotation and phylogenetics, we simplify one critical step in the pathogen analysis workflow.

**ANNOUNCEMENT**

The R programming language offers many powerful packages for phylogenetic analyses and visualization (e.g., ape (1) and ggtree (2)). Although R workflows are powerful, its command line interface can be difficult and time-consuming to master to create data visualizations. Many researchers might opt instead to create a figure using software with a graphical user interface (GUI). By leveraging Shiny (3), R’s native library for creating interactive applications, we can harness the power of R and its phylogenetic packages while abstracting away some of the programming complexity, making it easier to visualize and annotate phylogenetic trees for the nonexpert.

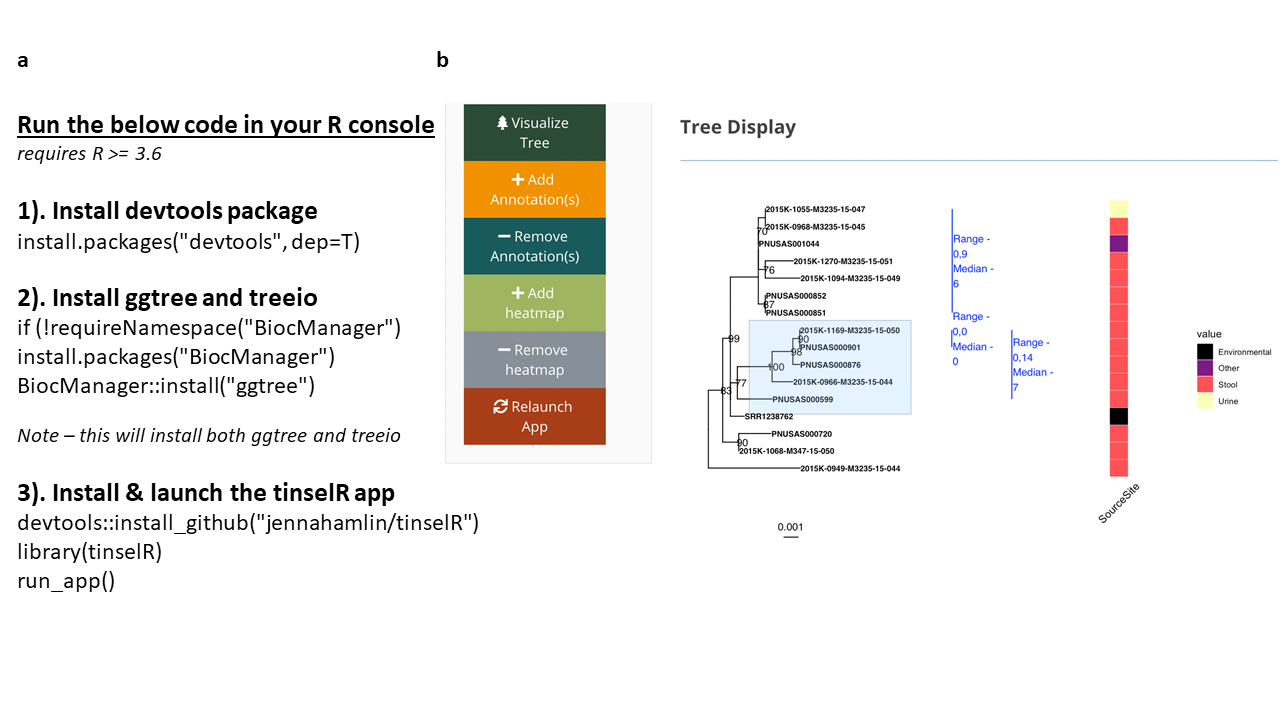
To this end, we developed **tinselR** (pronounced tinsel-er), a R Shiny application that can be run locally or deployed to the cloud. tinselR’s minimum input requirement is a Newick-formatted phylogenetic tree, but it can also take a genetic distance matrix of single nucleotide polymorphisms (SNPs) along with user-defined metadata. Once loaded, one can quickly modify the appearance of the plotted tree to include annotations, relabel tips, or add a heatmap. The modified tree images are downloadable as PDF, PNG, or TIFF for presentations, publications, or other communication with collaborators and stakeholders.

**INSTALLATION, EXAMPLE DATA, and ADDITIONAL RESOURCES**

To install tinselR from GitHub, users will need to install the R package devtools (4) and R version >= 3.6. The R packages ggtree (2) and treeio (5) are also required and can be installed from Bioconductor using BiocManager). With the installation of these dependencies, tinselR is installable via the install\_github command from devtools. Explicit installation commands are below (Figure 1a), and the final command (run\_app) will launch the application locally. Note that install\_github will also install other missing R dependencies. TinselR will accept Newick tree files from any program, e.g., RAxML (6), as input. Although it is possible to host R Shiny applications on a server, to date tinselR has only been tested by single users running the application locally. We recommend testing to ensure tinselR performs as expected under multi-user conditions before providing access from a server for production purposes.

After launching tinselR, new users can explore the application using one of the pre-loaded datasets located in the ‘Example Data’ tab. We provide three datasets (i.e., Newick formatted tree, genetic distance matrix, and metadata file) available once the application is launched. After clicking on the ‘Example Data’ tab, a user can select one of the datasets (e.g., example data 1, 2, and 3) from the drop-down menu to test the application. We highlight the capabilities of tinselR using example data 1 (Figure 1b). The example data are either *Escherichia coli* (from NCBI Bioproject: PRJNA218110) or *Salmonella enterica* (from NCBI Bioproject: PRJNA230403) with the number of isolates ranging from 14 - 19. The genomic data used in the example data sets were generated and used under the CDC IRB protocol 7172.

Additional documentation for tinselR is located at this site -  <https://jennahamlin.github.io/tinselR/>, specifically a vignette which describes the input data, example data, and a more detailed description for how to use the application. The source code and ability to file an issue is located at this site - <https://github.com/jennahamlin/tinselR>.



**Figure 1. a)** Code which will install and launch the tinselR application. Please visit the GitHub page to determine the release version for installation and specify that in the devtools::install\_github command or install the latest source using this command: devtools::install\_github("jennahamlin/tinselR@\*release") . **b)** Example dataset 1 displayed with SNP annotations and a heatmap indicating collection source.

**Acknowledgements**

We would like to thank those who participated in testing the application and providing valuable feedback during code review including the BioMe team at CDC. We also thank J. Notoma for helping to get tinselR up and running on the CDC internal server. This publication was supported by Cooperative Agreement Number 60OE000103, funded by Centers for Disease Control and Prevention through the Association of Public Health Laboratories. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of Centers for Disease Control and Prevention or the Association of Public Health Laboratories.

**References**

1. Paradis E, Schliep K. 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**: 526-528.
2. Yu G, Smith DK, Zhu H, Guan Y, Lam TT. 2017. Ggtree: An R Package for Visualization and Annotation of Phylogenetic Trees with Their Covariates and Other Associated Data. *Methods Ecol Evol* **8**: 28–36.
3. Hadley W, Hester J, Chang W. 2021. devtools: Tools to Make Developing R Packages Easier. R package version 2.4.0. <https://CRAN.R-project.org/package=devtools>
4. Chang W, Cheng J, Allaire JJ, Xie Y, McPherson J. 2017. Shiny: Web application framework for R. <https://cran.r-project.org/web/packages/shiny/index.html>
5. Wang LG, Lam TT, Xu S, Dai Z, Zhou L, Feng T, Guo P, Dunn CW, Jones BR, Bradley T, Zhu H. 2020. Treeio: An R Package for Phylogenetic Tree Input and Output with Richly Annotated and Associated Data. *Mol Biol Evol* **37**: 599–603.
6. Stamatakis A. 2014. RAxML Version 8: A Tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies. *Bioinformatics* **30**: 1312–3.