

TinselR – An RShiny Application for Annotating Outbreak Trees

Running Title: TinselR for annotating phylogenetic trees

Jennafer A. P. Hamlin^{current,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

^{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 designed to simplify this critical step in the pathogen analysis workflow.

ANNOUNCEMENT

The R programming language offers many powerful packages for phylogenetic analyses and visualization (e.g., ape (Paradis, Claude, and Strimmer 2004) and ggtree (Yu et al. 2017)). Although R workflows are powerful, its command line interface can be difficult and time-consuming to master to create a publication-ready figure. Many researchers might opt instead to create a figure using software with a graphical user interface (GUI). By leveraging Shiny (Chang et al. 2017), R's native library for creating web 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 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) and metadata. Once loaded, a user 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 (Wickham and Chang 2016). The R packages ggtree (Yu et al. 2017) and treeio (Wang et al. 2020) are also required and can be installed from Bioconductor using BiocManager (Morgan 2019). 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 (Stamatakis 2014), as input. Although it is possible to host RShiny 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). After clicking on the 'Example Data' tab, users 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 (Figure 1b) using example data 1. 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 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>.

a

Run the below code in your R console

1). Install devtools package

```
install.packages("devtools", dep=T)
```

2). Install ggtree and treeio

```
if (!requireNamespace("BiocManager"))  
install.packages("BiocManager")  
BiocManager::install("ggtree")
```

Note – this will install both ggtree and treeio

3). Install & launch the tinselR app

```
devtools::install_github("jennahamlin/tinselR")  
library(tinselR)  
run_app()
```

b

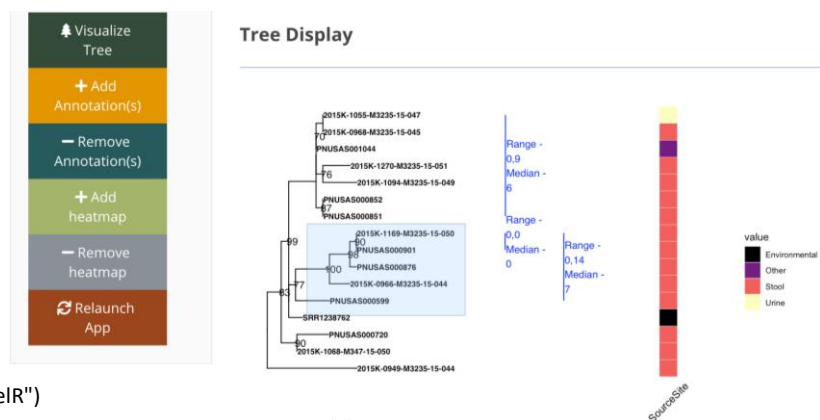


Figure 1. a) Code which will install and launch the tinselR application. Please visit the GitHub page to determine the release version for installation and specify that in the devtools::install_github comand. **b)** Example dataset 1 displayed with annotations and a heatmap indicating collection source.

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 gett 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

- Chang, Winston, Joe Cheng, J Allaire, Yihui Xie, Jonathan McPherson, and others. 2017. “Shiny: Web Application Framework for R.” *R Package Version 1* (5).
- Morgan, M. 2019. “BiocManager: Access the Bioconductor Project Package Repository. R Package Version 1.30. 10.”
- Paradis, Emmanuel, Julien Claude, and Korbinian Strimmer. 2004. “APE: Analyses of Phylogenetics and Evolution in R Language.” *Bioinformatics* 20 (2): 289–90.
- Stamatakis, Alexandros. 2014. “RAxML Version 8: A Tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies.” *Bioinformatics* 30 (9): 1312–3.
- Wang, Li-Gen, Tommy Tsan-Yuk Lam, Shuangbin Xu, Zehan Dai, Lang Zhou, Tingze Feng, Pingfan Guo, et al. 2020. “Treeio: An R Package for Phylogenetic Tree Input and Output with Richly Annotated and Associated Data.” *Molecular Biology and Evolution* 37 (2): 599–603.
- Wickham, Hadley, and Winston Chang. 2016. “Devtools: Tools to Make Developing R Packages Easier.” *R Package Version 1* (0): 9000.
- Yu, Guangchuang, David K Smith, Huachen Zhu, Yi Guan, and Tommy Tsan-Yuk Lam. 2017. “Ggtree: An R Package for Visualization and Annotation of Phylogenetic Trees with Their Covariates and Other Associated Data.” *Methods in Ecology and Evolution* 8 (1): 28–36.