treesifter

Introduction

treesiftr is a Shiny application (Chang et al. 2018) for visualizing the relationship between phylogenetic trees and phylogenetic data. Phylogenetic trees are crucial to the study of comparative biology, taxonomy, and evolution. However, understanding how to read a phylogenetic tree, and how a phylogenetic tree relates to underlying phylogenetic data, remains challenging.

In today's lab exercise, we will learn about a phylogenetic matrix, and then use the data in the matrix to visualize a phylogenetic tree.

> *Tip*
>
> Linked text goes to the glossary. If you see a term you don't recognize, remember you can refresh your memory at the bottom of this worksheet!

The fossil bear matrix

The data matrix included with treesiftr is a matrix of binary ("0" and "1") characters compiled to estimate a topology of living and extinct bear species (Abella et al. 2012). This matrix is fairly typical in size for a paleontological matrix, comprising 62 characters. It is, however, atypically complete, with only 18% missing data. In the following exercises, missing data will be represented by a thin black line. The "0" state will be represented in pale blue, and the "1" in brown.

treesiftr

Navigate to the treesiftr application.

treesiftr works by subsetting a phylogenetic matrix using the **start** and **step** arguments. These are found on the left-hand side of the screen. The **start** argument controls where in the matrix you would like to begin visualizing characters. For example, a start value of 1 would indicate to begin visualizing characters from the first character in the matrix. The **step** value indicates how many characters at once to visualize. A step value of three would indicate characters should be viewed in threes. For example, if **start** = 1 and step = **3**, the visualization will show characters 1, 2, and 3. The start and end points of character visualizations are noted in the upper left-hand corner of the visualization.

There are also two checkboxes. Leave these unchecked until otherwise noted.

treesiftr works by subsetting a phylogenetic matrix by user input. Then, a parsimony tree is constructed in Phangorn (Schliep 2011, Schliep2017) from the user-defined subset. The tree is scored under both parsimony and Lewis' Mk model (Lewis 2001) for discrete character data. The data and tree are then visualized using ggtree (Yu et al. 2017), based upon the ggplot2 package (Wickham 2016). This application makes use of Shiny to provide a graphical interface, but there is a second included tutorial for more experienced users of the R statistical language.

Parsimony Questions

1. Visualize characters 1 and 2. What is the parsimony score for this character set. Click "Do you want to print the parsimony score?" in the interface to check your answer.

2. Visualize characters 2 and 3. What monophyletic group from tree 1 is no longer on this tree? Click below for the answer

3. What is the parsimony score of the 31-34 character set? Click "Do you want to print the parsimony score?" in the infterface to check your answer.

4. Which character, 8, 9 or 10, represents a reversal?

5. What information would we need to decide if the "1" state possesed by *Zaragocyon_daamsi* in character 52 is an autapomorphy?

6. Click the switch that says "Do you want to print the likelihood score under the Mk model?" *and* the switch that says "Do you want to print the parsimony score?". Do all characters with the same parsimony score have the same likelihood score?

7. Compare characters 46-49 and 47-50. Why does set 47-50 have a better likelihood than 46-49?

8. What is the relationship between the likelihood score and increasing the number of characters visualized?

9. What is the minimum number adding a character can add to the parsimony score?

10. These trees are fully resolved. Based on your exploration of the data, does this degree of resolution make sense?

Glossary

Ancestral State: A character state possessed by the ancestor of a group

Autapomorphy: A character state that is unique to a specific taxon.

Derived State: A character state that is different from the ancestral state.

Likelihood Score: The likelihood of the observed data under a specific model.

Maximum likelihood: A phylogenetic optimatlity criterion under which phylogenetic data are modeled according to sets of assumptions. Under this criterion, the tree that has the best ("maximum") likelihood score under the assumed model is to be preferred.

Maximum parsimony: A phylogenetic optimality criterion. This criterion holds that the tree implying the fewest changes in the characters used to generate it should be preferred.

Monophyletic: A group on a phylogeny of an ancestor and all of its descendents.

Parsimony Score: The number of changes implied by a character on a tree.

Reversal: A change from the derived state back to the ancestral state.

References

Abella, Juan, David M. Alba, Josep M. Robles, Alberto Valenciano, Cheyenn Rotgers, Raül Carmona, Plinio Montoya, and Jorge Morales. 2012. "*Kretzoiarctos* Gen. Nov., the Oldest Member of the Giant Panda Clade." *PLoS One* 17: e48985.

Chang, Winston, Joe Cheng, JJ Allaire, Yihui Xie, and Jonathan McPherson. 2018. *Shiny: Web Application Framework for R*. https://CRAN.R-project.org/package=shiny.

Lewis, Paul O. 2001. "A Likelihood Approach to Estimating Phylogeny from Discrete Morphological Character Data." *Systematic Biology* 50 (6). Oxford University Press: 913–25.

Schliep, K.P. 2011. "Phangorn: Phylogenetic Analysis in R." *Bioinformatics* 27 (4): 592–93. https://doi.org/10.1093/bioinformatics/btq706.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis.* Springer-Verlag New York. http://ggplot2.org.

Yu, Guangchuang, David 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. https://doi.org/10.1111/2041-210X.12628.