Constantinos Tsirogiannis
Aarhus University

# Fast Computations of Phylogenetic Diversity Measures

The Problem

During field trips, ecologists record all species of a specific family that live at a certain location. One of the main questions that arise during this process is: are those species closely related to each other?

Approach:
Compute Distances on a Phylogenetic Tree

To calculate how closely related are these species, biologists use a phylogenetic tree; the leaf nodes on this tree represent species that exist today, the interior nodes represent possible ancestor species, and the links between the nodes have a distance value (e.g. time since last speciation) . On this tree, biologists pick X leaf nodes that represent the species they want to study (with red colour in the figure), and then they compute a distance measure; for example the average path cost between any pair of red leaves.
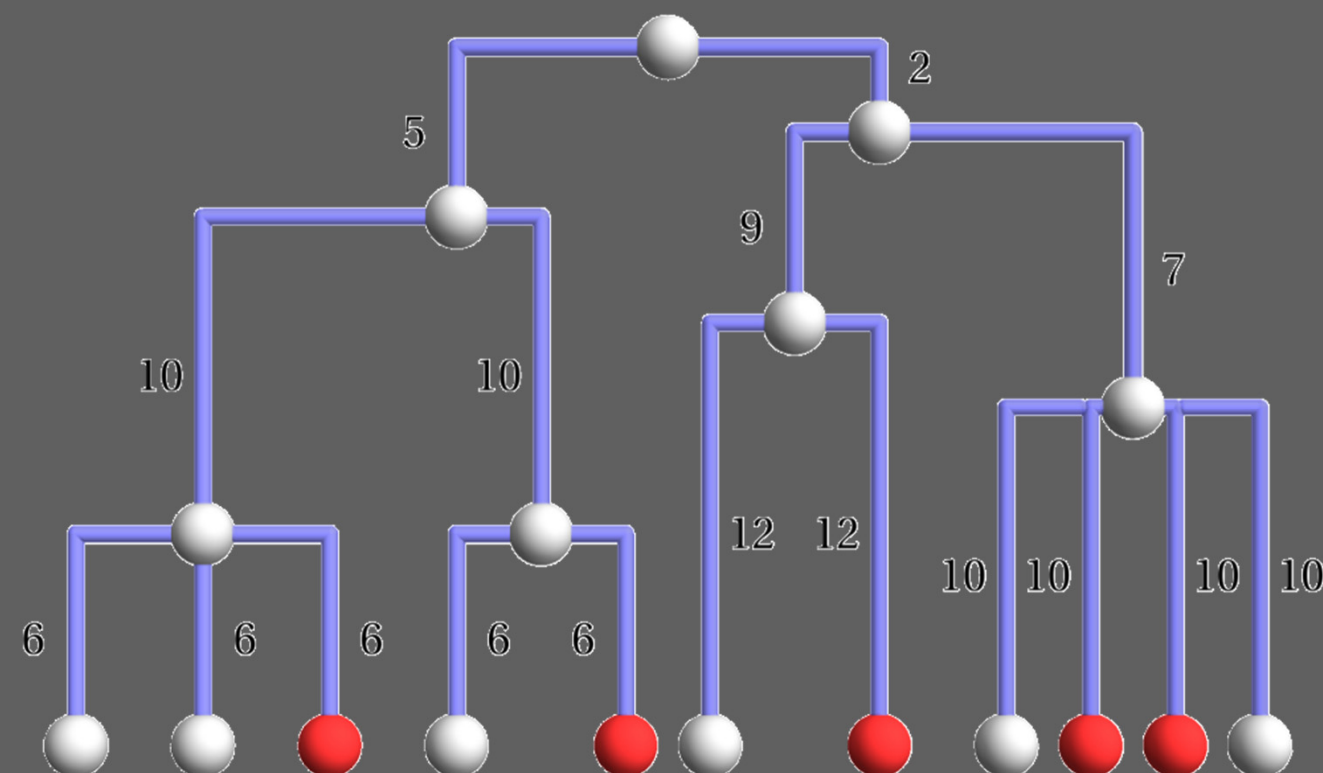
Calculating Measure Statistics

But, computing a distance value is not enough. For instance, suppose that the distance value that we calculated for the selected leaves is 42. To answer our initial question ("are these species closely related?") we should know if 42 is relatively large or small. In other words, we should know how 42 compares with the distance value that we would get if we picked at random X species in the tree.

Optimally, we would like to know the mean, variance and other statistics of the distance value that we want to compute when choosing at random X species from the same tree.

Unfortunately, until recently there were no available algorithms or software for solving this problem.

## Our Contribution: a New Software Package
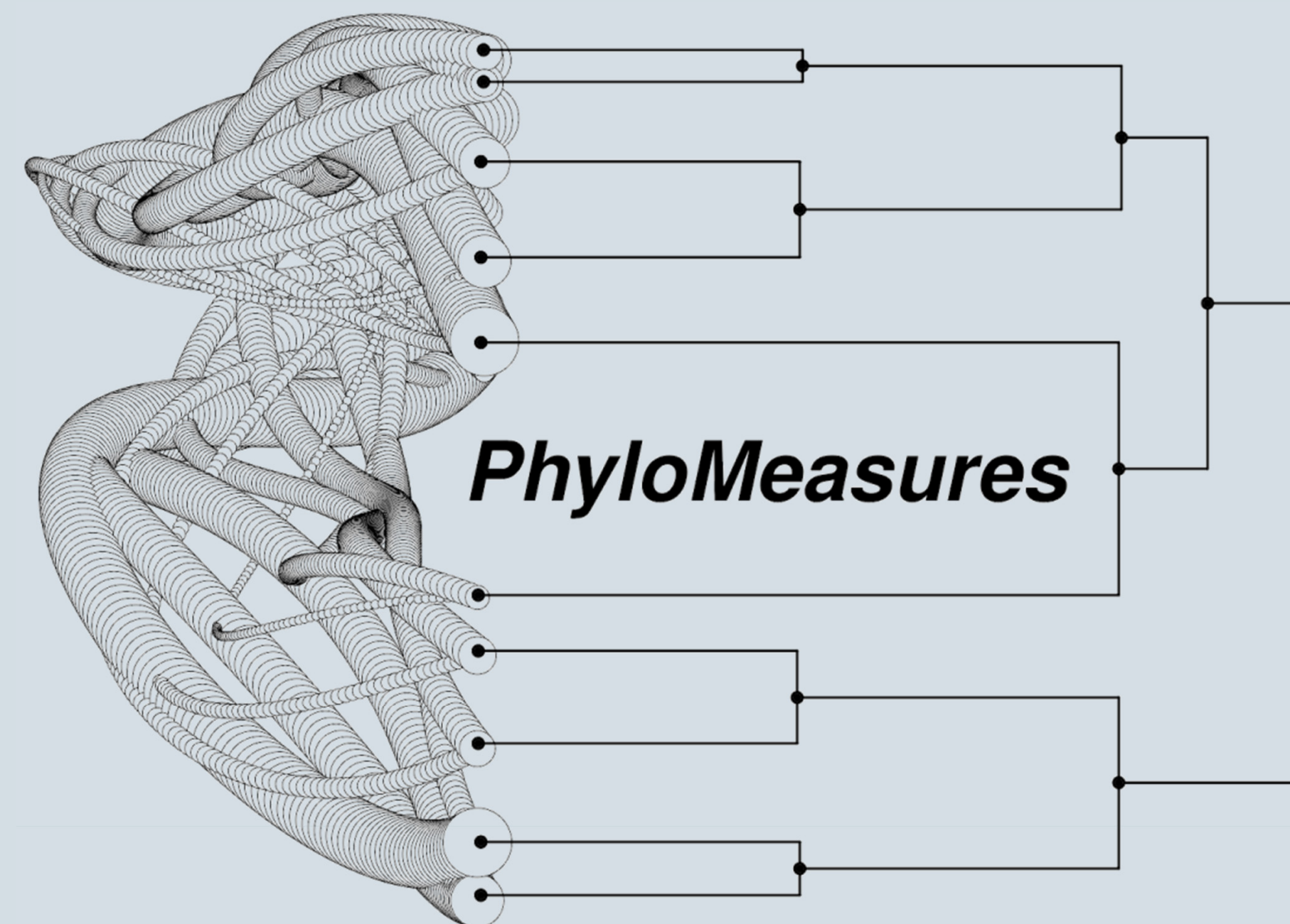
PhyloMeasures

We developed the open source software package PhyloMeasures which provides efficient programs for computing phylogenetic distance measures and their statistics.

The programs of this package are based on efficient algorithms that we introduced in a series of publications during the last three years. As a result, the package programs have a very fast performance even on trees that have tens of thousands of leaves.

PhyloMeasures is available under the GNU public license 3, and comes both as a C++ library and an R package.

Comparison with other Software

We measured the running time of PhyloMeasures on trees of very large sizes, and we compared that with the performance of the most popular similar package so far, the R package picante. Our package is much more efficient; for certain phylogenetic measures our package is more than 300,000 times faster even on medium size trees.
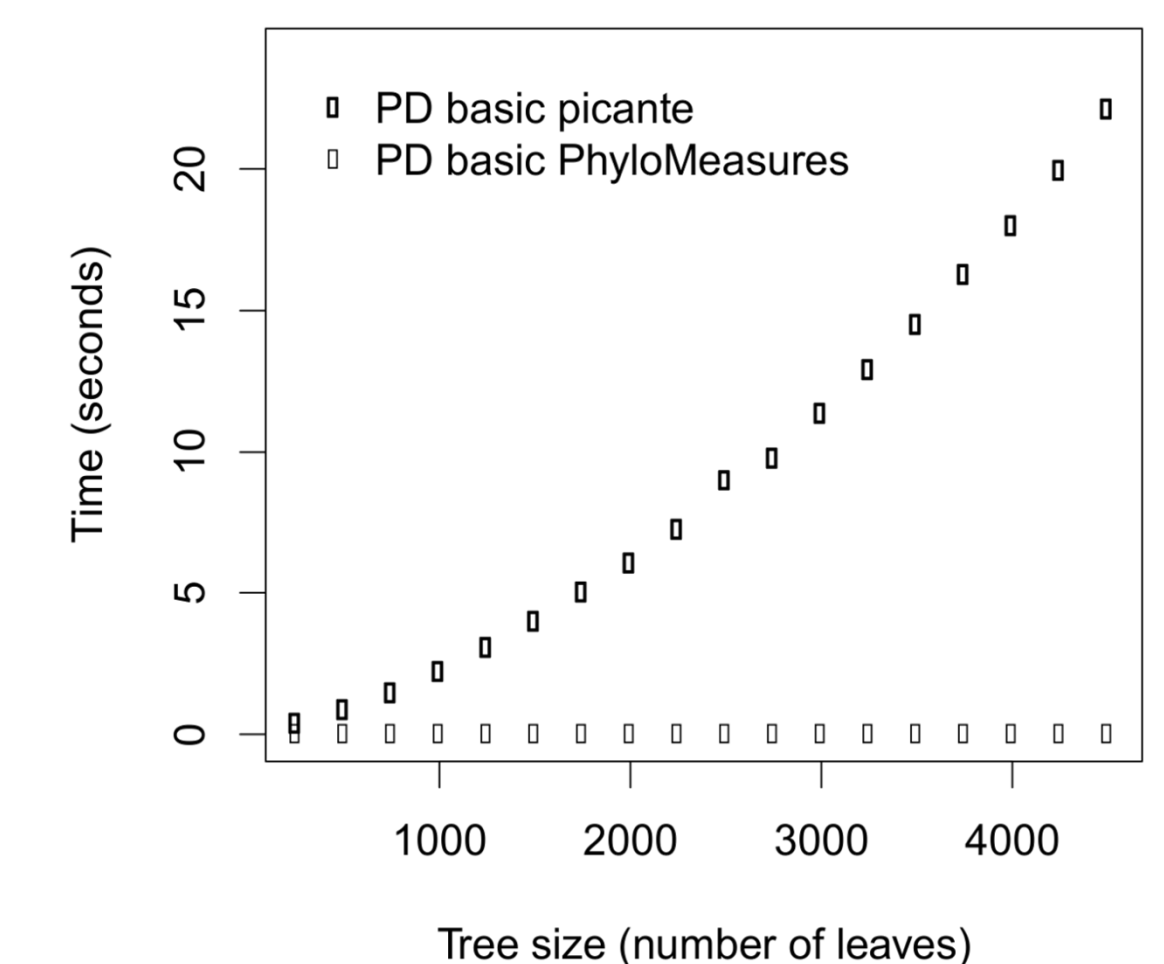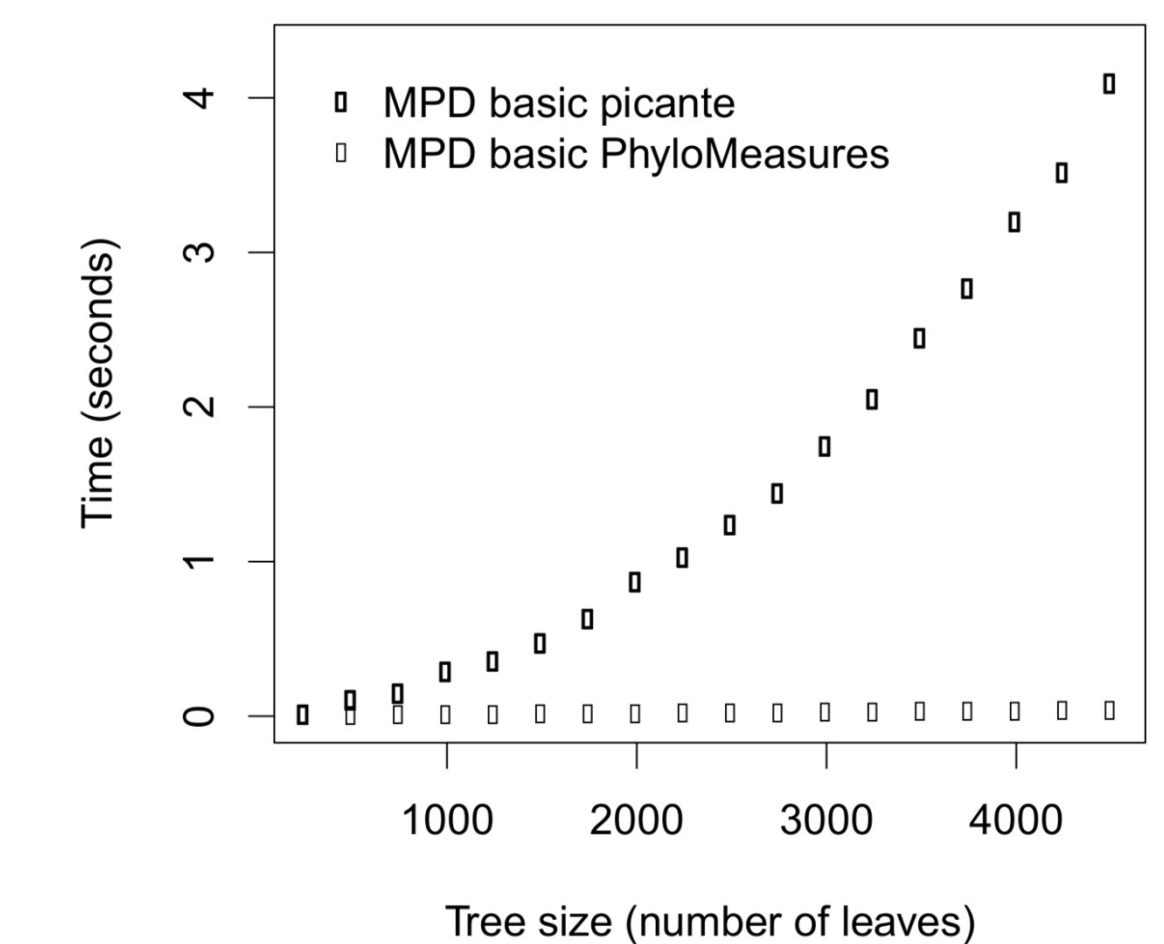
References

[1] C. Tsirogiannis and B. Sandel. *PhyloMeasures: A Package for Computing Phylogenetic Biodiversity Measures and Their Statistical Moments*. Manuscript.

[2] phylomeasures.blogspot.com