

# A WEB-BASED COLLABORATIVE PLATFORM FOR COMPARING PHYLOGENIES

Nicolas FIORINI, Antoine BISCH, Florent DUMOND, Mawusse AGBESSI, François CHEVENET, Vincent LEFORT, Anne-Muriel CHIFOLLEAU and Vincent BERRY

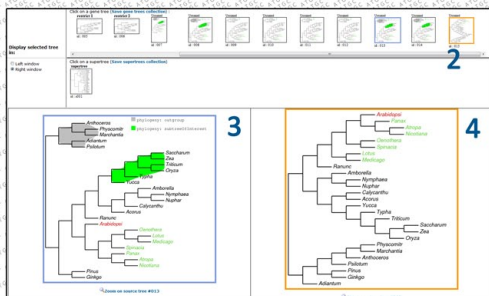
<http://www.atgc-montpellier.fr/compiphy>

[vberry@lirmm.fr](mailto:vberry@lirmm.fr)

## Objectives

Comparing phylogenies is a routine process in bioinformatics, performed to analyze alternative trees for a sequence alignment, to detect an outlier tree in a collection of trees or to view differences between gene and species trees. This in turn allows to identify recombinant evolution such as horizontal gene transfers or to evaluate the ability of a supertree/supermatrix tree to summarize a collection of individual gene trees.

To locate (dis)similar topologic features of analyzed trees, it is often useful to color particular leaves. When comparing trees with differing taxa sets, it helps to temporarily restrict these trees to their common taxa. Highlighting their common topologic structure is also important when they disagree on the position of few taxa [1]. Moreover, during the course of the analysis, the content of the tree collection or the names of the trees usually need to be modified to reflect the analysis progress. It is also useful to attach documents to single trees such as alignments or description files, or to the whole collection such as task lists or a summary of the performed analysis.



## Tree comparison

CompPhy allows to visualise simultaneously two selected trees (gene trees and/or supertrees) and to highlight their common parts in order to compare them.

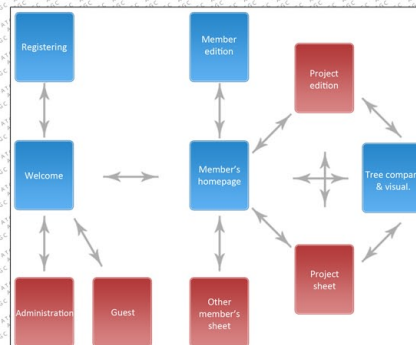
## CompPhy use case diagram

This diagram presents an overview of the functionalities provided by CompPhy for each kind of user (classical user, project manager and guest). The green boxes correspond to the functionalities already implemented. The green (resp. red) boxes show the functionalities that are already implemented (resp. that will be implemented soon).



## CompPhy tools and options

Figure 4 shows a screenshot of the CompPhy tools and options interface. It displays various settings and options for users, including tree selection, comparison parameters, and visualization options.

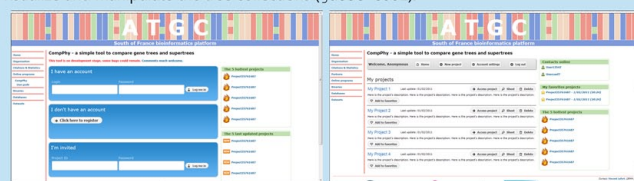


## Navigating between web pages

This figure presents the various pages of the CompPhy web site and the possibilities to navigate among them. The blue boxes correspond to the pages already implemented.

## A typical session in CompPhy

- (1) A PhD student S works on horizontal transfers (HGT) in bacteria, in collaboration with a biologist B and a computer scientist C, the three of them being in different labs. Prior to a meeting by videoconference, S loads his gene trees, their respective alignment and a tree M computed by a supertree approach (project creation/load trees/load additional files).
- (2) Each participant to the distant meeting then connects both to the videoconference tool and with a web browser to the CompPhy project, centralizing the data to be analyzed.
- (3) S presents his first conclusions on the presence of HGTs by displaying on the shared visualization page the gene trees (zone 3) and the comprehensive species tree (zone 4), highlighting in the process the species suspected to be involved in HGTs (colorize tool).
- (4) B advocates that part of S conclusions are wrong due to a species group G being misplaced in M with respect to current taxonomy knowledge. To know if this results from gene amalgamation, C uses the supertree approach to measure the degree of conflicting signal in the gene tree collection (supertree tool). The supertree is globally well resolved, and in accordance to current knowledge. This validates the collection of gene trees as a whole.
- (5) Then, B and S examine the trees (zones 2, 3 and 4) and eventually identify a gene tree T having the same problem as M (restrict tree and MAST tools). By looking at the alignment A corresponding to T, they see that it contains more sites than other genes, thus has a great influence when producing M (additional files & alignment visualization tools). A was built automatically by S and is badly aligned due to some indels. B then corrects the alignment by hand (download & upload additional files) and then asks S to recompute his concatenate matrix to obtain a new comprehensive tree M (download files & upload trees & change tree names tools).
- (6) Eventually, they end up with a colorization of the gene trees putting the transferred part into evidence, and a picture of the new comprehensive tree M where each species name is annotated in the margin by two histograms indicating the relative number of time the species was identified to be a donor, resp. a receiver of an HGT (manual tuning tool).
- (7) They submit a paper detailing their findings to a journal whose referees can anonymously visualize and manipulate the tree collections (guest tool).



## Perspectives

The platform is still under development (<http://compiphy.creatox.com>) and collaborative operations will be added such as discussion threads and the possibility to invite guests, including anonymous reviewers for trees appearing in a paper submitted to a journal. Advanced highlighting operations will also be proposed such as indicating the largest common substructure of selected trees (MAST) or automatic coloring of subtrees based on taxonomic information.

To evaluate comprehensive trees (i.e., supertrees or supermatrix trees), several tools will be proposed such as the identification of arbitrary branches with respect to a collection of gene source trees (PhysIC method), or such as the computation of support measures for branches in a comprehensive tree depending on a selection of source trees. Tree comparison will be eased by functionalities such as masking whole subtrees, e.g. replacing them by expandable nodes; swapping subtrees of a same node to maximise the horizontal correspondence between taxa of the two displayed trees.

## References & Acknowledgment

We thank researchers from the ISEM and University Montpellier 1 for their guidance in the realization of this tool. This work is supported by the phylariane ANR-08-EMER-011-01 and the Région LR.

- [1] T. Nye, P. Lio, W. Gilks. A novel algorithm and web-based tool for comparing two alternative phylogenetic trees. *Bioinformatics*, 22:117-119, 2006 (website <http://www.mas.ncl.ac.uk/cgi-bin/ntrm/pairwise.cgi>).
- [2] F. Chevenet, O. Croce, M. Hebrard, R. Christen, V. Berry (2010). SripTree: scripting phylogenetic graphics. *Bioinformatics*, 26, 8:1125-6, 2011.

## CompPhy: an online toolbox

The CompPhy platform allows a group of distant coworkers to upload and manage a collection of phylogenetic trees into a project environment. CompPhy offers a visual comparison of two trees chosen in the project's collection of source trees and supertrees.

In order to perform efficient comparisons, CompPhy implements a collection of tools able to perform:

- taxa colorization to highlight specific species among various trees
- tree restriction to common taxa to see more clearly the agreements and disagreements between selected sources trees
- supertree computation for selected source trees.

As CompPhy relies on the SripTree system [2] to display trees, complex highlighting and annotation operations can be performed manually for each tree or supertree.