

Which plants would you choose to study for new drug discovery? A pipeline for a phylogenetic approach

Elisa Toini^{1,2}, Giovanni Zecca^{1,2}, Maria Totaro^{1,3}, Massimo Labra^{1,2}, Fabrizio Grassi^{1,2}

E-mail: e.toini@campus.unimib.it

¹ Department of Biotechnology and Biosciences, University of Milano-Bicocca, Piazza Della Scienza 2, 20126 Milan, Italy

² NBFC, National Biodiversity Future Center, 90133 Palermo, Italy

³ Faculty of Education, Free University of Bozen-Bolzano, Brixen-Bressanone, Italy

Keywords: phylogenetic methods, medicinal plant, drug discovery

Abstract:

Some plants contain useful molecules for human health but identifying them requires time and resources. To optimize the discovery of new useful compounds, a strategic approach is needed to first select plants with high potential. One strategy is the phylogenetic approach, as closely related plant species tend to share biochemistry and medicinal properties. The aim is to propose a pipeline of phylogenetic methods to identify plants most likely to contain beneficial molecules.

Thus, a phylogenetic tree was downloaded from the literature, and five monophyletic subtrees were extracted from it, containing a total of 30,388 plant species. Then, lists of medicinal plants correlated with 12 diseases were obtained from the CMAUP medicinal plants database, and lists of medicinal plants correlated with 12 biological activities were obtained from Dr. Duke's Phytochemical and Ethnobotanical Databases.

The data from each list were combined with the data from each monophyletic subtree, and for all of them, the phylogenetic signal was measured using different methods. In the presence of phylogenetic signals, subsequent analyses were done to determine the exact position of phylogenetic clumping by identifying the “hot nodes”, nodes related to the plants with the most potential. The trees containing the species descending from the hot nodes were extracted, and those with more than 14 tips were plotted. Then, for all trees with more than 100 tips, another method called “hidden state prediction” was applied to calculate the probability of a plant on the tree having medicinal properties.

From the analysis of the 12 lists, a total of 198 trees were plotted. On each tree, the plants descending from the most hot nodes and with the highest probability are the plants with the highest potential to have medicinal properties. This work was able to apply a pipeline of different phylogenetic methods that can be used for the selection of potential plants for drug discovery; the functions used in the pipeline will soon be available in the R package “pm4mp”.