Geophysical Research Abstracts
Vol. 20, EGU2018-6427, 2018
EGU General Assembly 2018

Assessing among-lineage variability in phylogenetic imputation of functional trait datasets

Rafael Molina-Venegas (1,6), Juan Carlos Moreno-Saiz (2), Isabel Castro Parga (3), Jonathan Davies (4), Pedro Peres-Neto (5), and Miguel Ángel Rodríguez (6)

(1) University of Bern, Institute of Plant Sciences, Biology, Switzerland (rmolina@us.es), (2) Departamento de Biología (Botánica), Universidad Autónoma de Madrid, 28049 Madrid, Spain, (3) Departamento de Ecología, Universidad Autónoma de Madrid, 28049 Madrid, Spain, (4) Department of Biology, McGill University, 1205 ave Docteur Penfield, Montreal, Quebec H3A1B1 Canada, (5) Department of Biology, Concordia University, 7141 Sherbrooke Street West, Montreal, Quebec H4B1R6 Canada, (6) Departamento de Ciencias de la Vida, Universidad de Alcalá, 28871 Madrid, Spain

Plant traits are commonly used to address many ecological and evolutionary questions. However, gathering trait information is often challenging because the collection of functional trait data is a time- and resource-consuming task. Indeed, even one of the largest and most comprehensive functional trait databases compiled to date (i.e. the TRY Plant Trait Database, Kattge et al. 2011) is highly sparse and incomplete for many species, and thus plant trait-based studies often deal with missing data. Phylogenetic imputation has recently emerged as a potentially powerful tool for predicting missing data in functional traits datasets. As such, understanding the limitations of phylogenetic modelling in predicting trait values is critical if we are to use them in subsequent analyses. Previous studies have focused on the relationship between phylogenetic signal and clade-level prediction accuracy, yet variability in prediction accuracy among individual tips of phylogenies remains largely unexplored. Here, we used simulations of trait evolution along the branches of phylogenetic trees to show how the accuracy of phylogenetic imputations is influenced by the combined effects of (1) the amount of phylogenetic signal in the traits and (2) the branch length of the tips to be imputed. Specifically, we conducted cross-validation trials to estimate the variability in prediction accuracy among individual tips on the phylogenies (hereafter "tip-level accuracy"). We found that under a Brownian motion model of evolution (BM, Pagel's $\lambda = 1$), tip-level accuracy rapidly decreased with increasing tip branch-lengths, and only tips of approximately 10% or less of the total height of the trees showed consistently accurate predictions (i.e. cross-validation R-squared > 0.75). When phylogenetic signal was weak, the effect of tip branch-length was reduced, becoming negligible for traits simulated with $\lambda < 0.7$, where accuracy was in any case low. Our study shows that variability in prediction accuracy among individual tips of the phylogeny should be considered when evaluating the reliability of phylogenetically imputed trait values. To address this challenge, we describe a Monte Carlo-based method that allows one to estimate the expected tip-level accuracy of phylogenetic predictions for continuous traits. Our approach identifies gaps in functional trait datasets for which phylogenetic imputation performs poorly, and will help ecologists to design more efficient trait collection campaigns by focusing resources on lineages whose trait values are more uncertain.