

EFFECTS OF TAXON SAMPLING ON PHYLOGENETIC STABILITY AND SPECIATION RATES. A. A. Superson¹, M. Culver², A. Spagnuolo², and F. U. Battistuzzi¹, ¹Oakland University, Department of Biological Sciences, ²Oakland University, Department of Mathematics and Statistics.

Introduction: Continued technological advancements for obtaining sequencing data has lifted constraints in the ability to reconstruct the Tree of Life (TOL) by providing genomic data for thousands of species that were not previously available. Unfortunately, despite the increase in the amount of data available to unite shared characters of life in one tree, various biological and analytical factors confound reconstruction attempts resulting in conflicting phylogenetic placement of many taxa. However, a resolved TOL and the placement of specific speciation events within the tree are of crucial importance to reach a greater understanding of the origin of life, the evolutionary trajectory of major species, and the resulting adaptations life has made in approximately four billions of years of evolution.

In particular, two aspects of a phylogeny should be considered: topology (ancestor-descendant relationship) and patterns of speciation rates (fast, constant, or slow speciation rates through time). Topology controversies for taxonomic relationships, especially within the backbone, limit the ability to use TOL to understand patterns of coevolution between biotic and abiotic factors on a planetary scale. This effect is most profound within Prokaryotes whose relative species abundance is under represented within the sequence data available and whose evolutionary history goes back much further than most Eukaryotes. Recent TOL reconstruction studies have shown phyletic discrepancies within topologies of Terrabacteria which have particularly strong implications for the timing of the colonization of land, and major adaptations of life such as oxygenic photosynthesis and resistance to environmental hazard. Information on patterns of speciation rates is, instead, virtually absent.

Methods: We utilized a dataset of 766 fully-sequenced proteomes from six phyla that compose the Terrabacteria superphylum. Current TOL reconstruction studies have exposed conflicting phylogenies for this superphylum, particularly in the placement of the Deinococcus-Thermus (DT) phylum, possibly caused by the DT species representation that is much lower compared to that of other phyla. To investigate this, we created a pipeline that allowed us to perform various permutations on this dataset to determine how altered sampling scenarios affect the accuracy of phylogenetic reconstruction and speciation rates.

Each sampling scenario provided a simple framework for analyzing diversification patterns and rates

within prokaryotes. For patterns, we used a discrete Robinson-Foulds metric to quantify the level of discord among permuted maximum-likelihood trees while for rates we utilized ordinary differential equations to model speciation events in timetrees. The model assumes evolution as a continuous process with speciation rates derived the number of nodes in a phylogeny for a given elapsed time. Changes in speciation rates can therefore be quantified by comparisons of cumulative number of nodes at each point in the timetree.

Results: Our empirical data shows that different taxon samplings affect phylogenetic reconstruction suggesting that to obtain a stable and accurate TOL sequencing efforts should be more evenly distributed across taxonomic categories. In particular, we see that only some of the possible topological configurations for a tree of five phyla are recovered by our permutation scenarios but the frequency of the ones obtained changes dramatically with the number and identity of the species selected. Preliminary results of the speciation model are showing constant evolutionary trends for prokaryotes in general and also for Terrabacteria in particular. Further investigation using permutations will determine if this trend changes with different species compositions.