Exploration of the High Likelihood Set of Phylogenetic Tree Topologies in a Systematic Manner

Robert Grützmann*

Department of Mathematics, Technical University Dresden, 01062 Dresden, Germany

Introduction

Because it frequently returns to the same tree topology, Bayesian Markov chain Monte Carlo explores tree space slowly. An alternative strategy would be to systematically explore tree space, never returning to the same topology. We present an efficient parallelized method for mapping out the high likelihood set of phylogenetic tree topologies via systematic search in this article, which we show to be a good approximation of the high posterior set of tree topologies on the data sets analysed. A topology's "likelihood" here refers to the tree likelihood for the corresponding tree with optimised branch lengths. This method is known as "phylogenetic topography" (PT) [1].

Description

While the fossil record of medusozoan cnidarians is limited, the group's origin is thought to have occurred near the end of the Ediacaran period, around 550-580 million years ago. Possible crown-group hydrozoans have been described from 500 million year old Cambrian deposits, implying that extant hydrozoans have an ancient origin that likely dates back to the period of rapid diversification of metazoan life, during which all major modern animal phyla emerged. Hydrozoans are of particular interest in the study of development evolution because their radiation gave rise to diverse life cycle strategies, coloniality and division of labour. This diversity is concentrated in the medusozoan clade Hydroidolina, which contains the vast majority of the 3,800 nominal hydrozoans species [2].

Determining statistical consistency (or inconsistency) of topology-based methods for inferring species trees 13-15, testing the multispecies coalescent mode, determining identifiability of species trees using linear invariants of functions of gene tree topology probabilities, delimiting species, designing simulation studies for species tree inference methods and inferring species trees are all applications of probabilities of gene tree topologies given species trees. We anticipate that similar applications will be useful for predicting gene tree topologies given species networks. It will be especially useful to be able to assess the performance of methods for inferring species trees in the presence of hybridization, as well as methods for inferring species networks. Knowing the distribution of gene tree topologies may also be useful for estimating the likelihood that two gene trees have the same topology, a quantity used in constructing the prior that models gene tree discordance in BUCKy, a programme commonly used to estimate species trees or concordance trees [3].

In this study, we benchmarked different single-copy marker genes (SCMs)

***Address for Correspondence**: Robert Grützmann, Department of Mathematics, Technical University Dresden, 01062 Dresden, Germany, E-mail: Robert_Gruetzman121@uniklinikum-dresden.de*

commonly used in multidomain bacterial and archaeal phylogenetics and we identified a set of SCMs that performs best for phylogenetic trees derived from concatenated alignments using recently developed tree certainty (TC) metrics. Furthermore, we compare different taxon sampling strategies and show how uneven phyla representation can have a significant impact on the resulting trees and lower their overall TC. We then reconstructed a high-resolution tree using the best-performing marker gene set and balanced taxon sampling across bacteria and archaea, which clarifies the phylogenetic relationships between several phyla and identifies several deep-branching nodes where the true topology remains unknown [4].

Because of the large phylogenetic distance between bacteria and archaea, only a few SCMs are suitable for inter-domain phylogenetic reconstruction. Nonetheless, several independent studies have discovered 30-40 orthologous protein families that can be used for this purpose, including RNA polymerase subunits, ribosomal proteins, tRNA synthetases and proteins implicated in intracellular trafficking. We first evaluated the occurrence of these SCMs in a curated set of 1,650 bacterial and archaeal genomes derived from the Genome Taxonomy Database, using a set of 41 SCMs that encompasses this set and has previously been used for this purpose.

Species fusion events the species tree is vertically oriented, "up" toward the root and "down" toward the leaves. Lineages enter from two branches directly below the internal node of the species tree, representing a species-merging event, from a coalescent backward-in-time perspective [5]. Based on an arbitrarily labelled diagram of species tree T, we label one of these branches "left" and the other "right." These labels are only used for bookkeeping; the labelling has no bearing on subsequent calculations. Lineages that enter the system from the left and right branches are referred to as "left inputs" and "right inputs," respectively. Each node x of T has exactly one branch that connects it to its immediate predecessor on T. The shared label is used to refer to this branch.

Conclusion

The PI approach was used to test the ability of the four sets to resolve their phylogenetic relationships in terms of sequence divergence. In the teleost phylogeny, the power peaks of the Total and Teleost sets were in the shallow range, whereas those of the Slow 1000 and Slow 500 sets were in the relatively deep range. As a result of this approach, the Slow 1000 and Slow 500 sets had greater power to resolve the basal relationships of teleost fish in terms of sequence divergence than the Total and Teleost sets.

Acknowledgement

None.

Conflict of Interest

No conflict of interest.

References

1. Albert, James S. and Alexandre Antonelli. "Society for the study of systematic

biology symposium: Frontiers in parametric biogeography." *Syst Biol* 66 (2017): 125-127.

2. Alexander, Helen K., Amaury Lambert and Tanja Stadler. "Quantifying age-dependent extinction from species phylogenies." *Syst Biol* 65 (2016): 35-50.

3. Alfaro, Michael E., Francesco Santini, Chad Brock and Hugo Alamillo, et al. "Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates." *Proc Natl Acad Sci* 106 (2009): 13410-13414.

4. Barido-Sottani, Joëlle, Timothy G. Vaughan and Tanja Stadler. "A multitype birth–death model for Bayesian inference of lineage-specific birth and death rates." *Syst Biol* 69 (2020): 973-986.

5. Beaulieu, Jeremy M. and Brian C. O'Meara. "Detecting hidden diversification shifts in models of trait-dependent speciation and extinction." *Syst Biol* 65 (2016): 583-601.

How to cite this article: Grützmann, Robert. "Exploration of the High Likelihood Set of Phylogenetic Tree Topologies in a Systematic Manner." J Generalized Lie Theory App 16 (2022): 350.