

Homoplasy matrices: a tool for choosing between alternative topologies?

Leonid Rusin¹

¹Department of Evolutionary Biochemistry, Belozersky Institute of Physico-Chemical Biology,
Moscow State University, 119992 Moscow – Russia
roussine@yandex.ru

Keywords. Phylogenetic inference, Homoplasy, Parsimony

Introduction

It is long known that indices of non-parametric bootstrapping significantly underestimate the probability of existence of a topological element, particularly, in cases when lineages analysed differ in rates of character substitution [1, 2]. Results of conventional bootstrap analysis often indicate low support for deeper nodes of the tree or suggest unresolved phylogenies. True topological elements present in some of equal trees are thus lost upon building a consensus. Here we present an easy method of estimating reliability of individual nodes of the tree and screening the equal tree space for reliable topological elements.

Some Premises

Parsimony is a fast and widely used approach of analysing character sets. However, it is also known to perform inconsistently under certain circumstances [3]. In particular, sharp disparities in rates of character substitution generate high contents of homoplasy in the data at the level of individual characters. If the amount of homoplasy is high enough, parsimony may graft taxa erroneously on the basis of homoplastic characters while searching for most parsimonious trees during additional sequence replicates. Ultimately, heuristic search may stall within a local tree island, which is isolated from the global optimum by branch swappings that can not be implemented given a current combination of taxa [4, 5]. Thus, reconstruction of a true topology is strongly related to the signal/noise ratio in the data.

The Method

We analysed an alignment containing complete SSU rDNA sequences from 36 nematode taxa and 6 metazoan groups. In nematodes sampled across the entire phylum the gene has clearly different rates of molecular evolution. Variable positions were gradually removed from the alignment according to the amount of change assigned to each position on the most parsimonious tree, which resulted in a series of 24 subalignments. Each subalignment was processed with parsimony (PAUP* 4.0beta10 package). Screening of equal trees over series of reconstructions revealed many topological elements, which also occur in the bootstrap consensus obtained on the basis of the initial alignment. However, composition of other nodes altered depending on the amount of variable positions removed. In order to assess reliability of an individual element, we computed values of the homoplasy index (HI) for each character assigned the synapomorphic status by the algorithm and screened for the least homoplastic ones. Analysis showed that HI values for characters, which support robust elements retained in the initial consensus are close to or equal zero. Search within subalignment topologies detected nodes not retained in the consensus and yet reconstructed on the basis of low homoplastic characters (HI close to zero). Some of them correspond to higher nematode taxa, which monophyly has been already substantiated with independent evidence (morphology or SSU rRNA secondary structure), and others were novel. However, some alternative topologies were supported by characters with almost equally high HI values. To estimate the total amount of homoplasy generated by these topologies, pairwise homoplasy distances were calculated for each tree inferred in a series of subsequent reconstructions. Homoplasy matrices thus obtained and visualised as “trees” give an idea of how much homoplasy is generated between taxa by a particular topology. Detailed analysis of homoplasy “trees” showed that lineages, which constitute robust clades in the initial consensus contain less amount of homoplasy with respect to each other and are thus situated close on tree representations of the matrices. This is by no means always the case for taxa in nodes reconstructed on the basis of highly homoplastic characters. Comparative analysis of actual phylogenies and the corresponding homoplasy “trees” showed that the composition of highly homoplastic nodes in most cases is in discord with the pattern of distribution of pairwise homoplasy distances, *i.e.* taxa joined by the algorithm generate high level of homoplasy in the

node and thus do not cluster together in homoplasy “trees”. Alternatively, some of local combinations of taxa did correspond to the pattern of homoplasy distance distribution and, thus, are likely to represent natural monophyletic clades.

We used the approach described to screen the tree space pooled over the series of reconstructions for topological elements that meet the minimum homoplasy requirement and compiled a phylogeny, which was not found in neither of heuristic searches. In order to test this phylogenetic hypothesis, we constrained combinations of taxa, which either were supported by low homoplastic characters or were in accord with corresponding homoplasy distance distribution patterns and used these constraints in parsimony analysis of the initial alignment. Heuristic approach failed to find a tree, which would fully coincide with the compiled phylogeny (a total of 23 equal trees found). It was compared with the found trees by the Shimodaira-Hasegawa one-tailed test using RELL bootstrap and received a maximum likelihood score.

Acknowledgements

The work was supported by grants RFBR 01-04-48832, 02-04-48958; RRSCCE 1712.2003.4; FEBS Collaborative Scholarship for Central & Eastern Europe.

References

- [1] J. H. Kim, General inconsistency conditions for maximum parsimony: effects of branch lengths and increasing numbers of taxa. *Systematic Biology*, 45: 363-374, 1996.
- [2] M. J. Sanderson and M. F. Wojciechowski, Improved bootstrap confidence limits in large-scale phylogenies, with an example from Neo-Astragalus (Leguminosae). *Systematic Biology*, 49:671-685, 2000.
- [3] J. Felsenstein, Cases in which parsimony and compatibility methods will be positively misleading. *Systematic Zoology*, 27: 401-410, 1978.
- [4] R. D. M. Page, On islands of trees and the efficacy of different methods of dranch swapping in finding most-parsimonious trees. *Systematic Biology*, 42 (2): 200-210, 1993.
- [5] D. L. Swofford, G. J. Olsen, P. J. Wadell and D. M. Hillis, Phylogenetic inference, in: D. M. Hillis *et al.* (eds.), *Molecular Systematics*, Sianuer Associates, 1996.