

Origins and losses of parasitism

an analysis of the phylogenetic tree of life with a parsimony-like algorithm

Abstract

Contents

| | | |
|-----|--------------------------|---|
| 1 | Introduction | 4 |
| 2 | Methods | 5 |
| 2.1 | Simulation | 5 |
| 2.2 | Implementation | 5 |
| 3 | Results | 6 |
| 4 | Discussion | 7 |
| | Bibliography | 8 |

1 Introduction

This paper is about the further development of parsimony algorithms for non-binary trees, applied to the currently largest phylogeny synthesis tree of Open Tree Of Life, with the application to the ancestral state reconstruction of parasitism.

Researchers of the phylogenies have been dealt with the ancestral state reconstruction in the 60s. The first methods were only brute force **TODO: Quelle, siehe Fitch: Camin and Sokal 1965**. Next came a set of parsimony algorithms such as: Fitch-parsimony [Fit71], Wagner-parsimony [SM87] ... **TODO: weitere?**.

With more and more data, there is now the possibility to use more information to calculate the probabilities of the ancestral states. In addition to the states of the leaves, algorithms could also use branch lengths. The likelihood based algorithms came more in interest.

Our focus came with another 'data extension'. We wanted to work with the biggest phylogenetic tree that exists at this moment, which goes over all observed species. For most **TODO: most?** species there is no phylogeny, but only a taxonomic classification. So the biggest 'phylogenetic tree' is a synthesis of phylogenetic trees filled with a taxonomic tree given by Open Tree of Life [HSA⁺15]. This tree is not binary and therefore the developed algorithms are not directly applicable.

In this work, we have looked at the algorithms that are generally suited to our data, to develop them further for the not binary case, and finally to compare their usability with our sythesis tree.

We have decided to consider only parsimony algorithms since we have no information on branch lengths and no other additional information like different transition probabilities of our states.

2 Methods

2.1 Simulation

- build random binary trees, tag these (parameters: parasites vs free-living, beta-distribution)
- run fitch-parsimony, wagner-parsimony, our parsimony like algorithm
- build not binary tree (poisson distribution?)
- run new algorithms
- compare trees (distances)

2.2 Implementation

3 Results

4 Discussion

Wie gut kommt unsere Simulation an die echte Datenlage heran.

Fehlerquote der Daten an sich?

Wie gut ist unsere Datenlage? 3 mio Knoten, 1.8 named species (leaf nodes), 200.000 leaf nodes mit Information.

Simulation von subtrees

Welche Teile des Baumes sind gut, an welchen muss noch viel geforscht werden.

Wieviele Origins haben wir gefunden, was bedeutet diese Zahl?

Parameter der Simulation:

- Wie ist die Verteilung der vergessenen internen Knoten? Zum Wurzelknoten hin mehr vergessen?
- Wie sehen die Übergangswahrscheinlichkeiten aus von P- \rightarrow FL und andersherum?
- Verteilung Parasiten zu Freilebend zu keine Information

Bibliography

- [Fit71] FITCH, Walter M.: Toward Defining the Course of Evolution: Minimum Change for a Specific Tree Topology. In: *Systematic Biology* 20 (1971), Nr. 4, 406-416. <http://dx.doi.org/10.1093/sysbio/20.4.406>. – DOI 10.1093/sysbio/20.4.406
- [HSA⁺15] HINCHLIFF, Cody E. ; SMITH, Stephen A. ; ALLMAN, James F. ; BURLEIGH, J. G. ; CHAUDHARY, Ruchi ; COGHILL, Lyndon M. ; CRANDALL, Keith A. ; DENG, Jiabin ; DREW, Bryan T. ; GAZIS, Romina ; GUDE, Karl ; HIBBETT, David S. ; KATZ, Laura A. ; LAUGHINGHOUSE, H. D. ; MCTAVISH, Emily J. ; MIDFORD, Peter E. ; OWEN, Christopher L. ; REE, Richard H. ; REES, Jonathan A. ; SOLTIS, Douglas E. ; WILLIAMS, Tiffani ; CRANSTON, Karen A.: Synthesis of phylogeny and taxonomy into a comprehensive tree of life. In: *Proceedings of the National Academy of Sciences* 112 (2015), Nr. 41, 12764-12769. <http://dx.doi.org/10.1073/pnas.1423041112>. – DOI 10.1073/pnas.1423041112
- [SM87] SWOFFORD, David L. ; MADDISON, Wayne P.: Reconstructing ancestral character states under Wagner parsimony. In: *Mathematical Biosciences* 87 (1987), Nr. 2, 199 - 229. [http://dx.doi.org/https://doi.org/10.1016/0025-5564\(87\)90074-5](http://dx.doi.org/https://doi.org/10.1016/0025-5564(87)90074-5). – DOI [https://doi.org/10.1016/0025-5564\(87\)90074-5](https://doi.org/10.1016/0025-5564(87)90074-5). – ISSN 0025-5564