The Shape of Phylogenies Under Phase-Type Distributed Times to Speciation and Extinction

Albert Ch. Soewongsono, Barbara R. Holland, Małgorzata M. O'Reilly
School of Natural Sciences (Mathematics)
albert.soewongsono@utas.edu.au

UNIVERSITY of TASMANIA AUSTRALIA

Introduction

- We develop a mathematical model using a general Coxian phase-type (PH) distribution in order to examine factors affecting tree balance and branch lengths of phylogenetic trees.
- Adequate understanding of the shapes of phylogenetic trees can help to explain the ecological diversity [3].
- Thus, a mathematical model is required to examine macro-evolutionary processes (speciation and extinction events) on trees. Early models assuming constant speciation and extinction rates fail to represent empirical trees [2].

Methods

We test the effect of the model parameters on tree balance and branch length (See Fig 1):

- We simulate 600 trees, each with 100 extant tips where (a) times to speciation follow a general Coxian PH distribution and times to extinction follow an exponential distribution, and (b) vice versa (See Fig 1).
- We (a) vary parameters for the speciation process and assume constant parameters for the extinction process, and (b) vice versa.
- We compute tree balance via the β statistic [1] and branch lengths via the γ statistic [4].

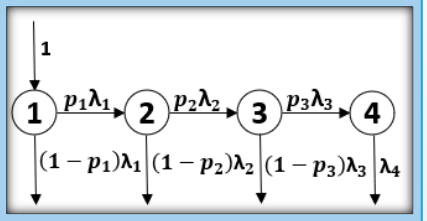
General Coxian PH Distribution

- A general Coxian PH distribution $T \sim PH(\boldsymbol{\alpha}, \boldsymbol{Q})$ models time T to absorption in a continuous-time Markov chain with transient states 1,…,n and an absorbing state 0.

Example:

- Initial distribution vector: $\boldsymbol{\alpha} = (1,0,0,0)$.
- Matrix $\boldsymbol{Q}$ of transition rates between transient states:

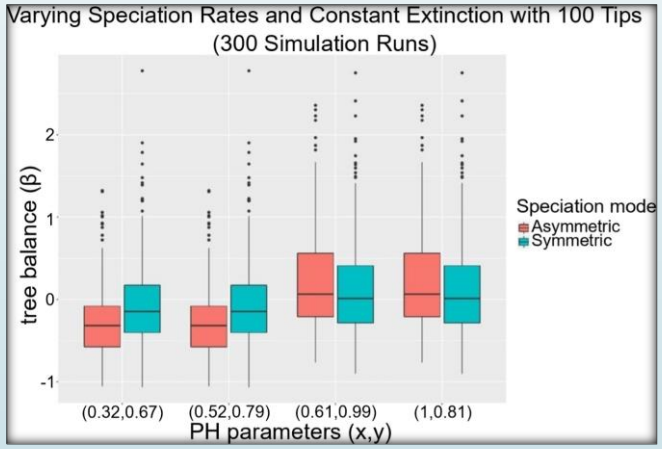
$$\mathbf{Q} = \begin{bmatrix} -\lambda_1 & p_1\lambda_1 & 0 & 0 \\ 0 & -\lambda_2 & p_2\lambda_2 & 0 \\ 0 & 0 & -\lambda_3 & p_3\lambda_3 \\ 0 & 0 & 0 & -\lambda_4 \end{bmatrix}$$



$$0 < p_1, p_2, p_3, p_4 \leq 1$$

Results

- We show that different parameter choices give trees that can range from highly balanced to highly unbalanced as measured by Aldous' β statistic.
- Our results suggest that tree balance is mainly controlled by speciation, rather than extinction.
- Our results also suggest that branch lengths are mostly controlled by extinction, rather than speciation.



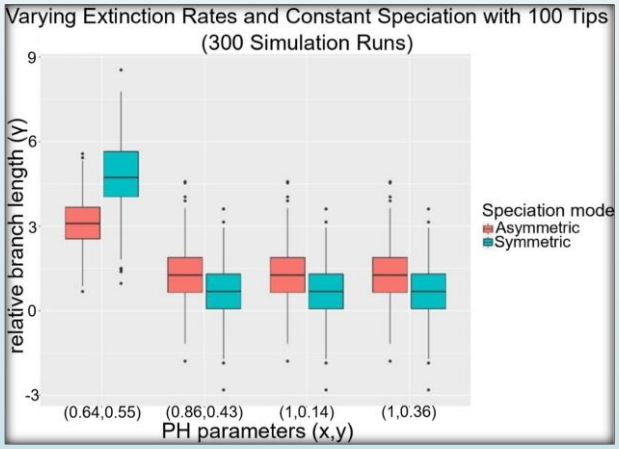
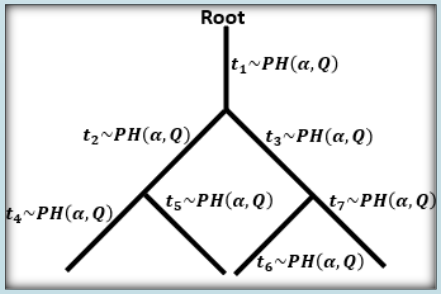


Fig 1. Effect of the model parameters on tree balance and branch lengths:
(a) speciation parameters affect tree balance (b) extinction parameters affect branch lengths



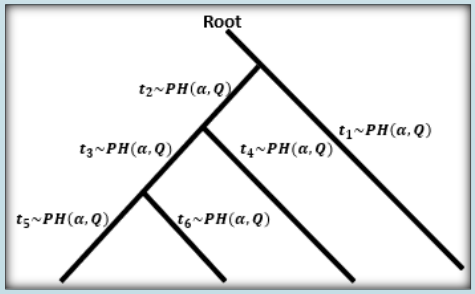


Fig 2. Balanced tree (left) and unbalanced tree (right) with 4 tips and PH-distributed branch lengths.

References

[1] Aldous, D. (1996). Probability distributions on cladograms. In Random discrete structures, pages 1–18. Springer.

[2] Aldous, D. J. et al. (2001). Stochastic models and descriptive statistics for phylogenetic trees, from yule to today. Statistical Science, 16(1):23–34.

[3] Mooers, A. O. and Heard, S. B. (1997). Inferring evolutionary process from phylogenetic tree shape. The quarterly review of Biology, 72(1):31–54.

[4] Pybus, O. G. and Harvey, P. H. (2000). Testing macro–evolutionary models using incomplete molecular phylogenies. Proceedings of the Royal Society of London. Series B: Biological Sciences, 267(1459):2267–2272.