

Causal inference and the detection of coevolutionary contingencies using dynamic phylogenetic models

E. Ringen, J. S. Martin, S. Claessens, & A. V. Jaeggi

2024-08-14

Abstract text here...

1 Introduction

Phylogenetic comparative methods (PCMs) are commonly used to study the co-evolution of organismal traits, spanning topics such as anatomy and physiology (Dunn and Ryan 2015; Garland, Bennett, and Rezende 2005; Navalón et al. 2019; O'Connor and Cornwallis 2022; Thayer et al. 2018), life history and behavior (Bielby et al. 2007; Clayton and Cotgreave 1994; MacLean et al. 2012; Salguero-Gómez et al. 2016), and cultural evolution (Mace and Holden 2005; Navarrete et al. 2016; Watts et al. 2016). “Coevolution” in its broadest sense refers to repeatable patterns of trait covariation over time, which can be investigated using a diverse family of statistical techniques depending on the research question and type of data available (Garamszegi 2014; Harvey and Pagel 1991; Nunn 2011). Many PCMs focus on understanding how traits coevolve by estimating evolutionary correlations and rates of change across phylogenetic trees, which can be used to better understand natural history and inform ancestral state reconstructions. However, evolutionary correlation does not imply evolutionary causation between traits (Shipley 2016). Thus, PCMs facilitating inference of why traits coevolve and change together across time are indispensable tools for testing macroevolutionary and ecological theory.

Among others, phylogenetic generalized linear (mixed) models (Grafen 1989; Hadfield and Nakagawa 2010; Symonds and Blomberg 2014), phylogenetic path analysis (Gonzalez-Voyer and Hardenberg 2014; Hardenberg and Gonzalez-Voyer 2013), and Pagel’s (1994) discrete method are the most popular approaches for assessing trait coevolution on macroevolutionary timescales. While each of these methods has clear benefits and performs well under specific scenarios, they are each limited in their generality by strong assumptions regarding the direction of causal effects among traits, the process of evolutionary change, and/or the statistical properties of the traits under investigation. As we detail further below, these modeling assumptions are likely to be violated in most datasets used for phylogenetic analysis, increasing the risk of inferential error and ultimately inhibiting our ability to explain the dynamics of phenotypic evolution. We therefore introduce a novel class of PCMs designed to address these challenges in a cohesive and flexible statistical framework, using Bayesian MCMC algorithms in the Stan statistical programming language (Carpenter et al. 2017).

We begin by briefly reviewing the benefits and constraints of current PCMs, particularly with regard to causal inference, exploring how their limitations motivate further methods development. We then formally and conceptually introduce a novel class of dynamic phylogenetic models. We provide a worked synthetic example of coevolution between two continuous traits – illustrated as female promiscuity and sperm length (Fitzpatrick et al. 2009) – to assess the accuracy and uncertainty of inferences made with dynamic phylogenetic models across a reasonable range of sample sizes for phylogenetic research. We also provide an accompanying coding tutorial using our **coevolve** R package to aid empiricists in easily applying basic dynamic phylogenetic models to their own datasets. We then demonstrate the generality and flexibility of our method with two empirical applications, which extend the model to complex, high-dimensional scenarios using dynamic latent

variables. In particular, we show how our method can generate insights into the causal dynamics of trait coevolution across both genetic and cultural evolutionary timescales, using (1) a comparative dataset on primate brain size, sociality, diet, and life history traits (DeCasien, Williams, and Higham 2017), which reduced to two coevolving latent variables, and (2) two studies on the evolution of social complexity across a global and a regional sample of pre-industrial human societies (Ringen, Martin, and Jaeggi 2021; Sheehan et al. 2023).

2 Current approaches and motivation for a novel method

Fundamental to PCMs is the adjustment of raw trait associations for shared evolutionary history using a phylogenetic tree (or set of trees) and a statistical model. In a basic sense, phylogenetic adjustment is crucial for facilitating causal inference, as shared evolutionary history tends to generate trait correlations among closely related species with similar phenotypes, creating the illusion of convergent co-evolution even when traits evolve independently. Adjustment for phylogeny reduces bias and variance due to clustering over the tree, thus reducing the risks of type I (false-positive), type II (false-negative), type M (magnitude), and type S (sign) errors during statistical inference (*CITE*). Nonetheless, adjustment for phylogeny is not a magic fix for all sources of unobserved confounding, nor does it guarantee that resulting estimates are causally interpretable (*CITE*). Therefore, while all PCMs employ statistical techniques that reduce bias, these methods vary widely in the degree to which they isolate the causal pathways by which traits coevolve over time, with most commonly used methods focusing largely on evolutionary correlation (Figure 1a) rather than causation (Figure 1b-e).

2.1 Phylogenetic generalized linear (mixed) models

2.2 Phylogenetic path analysis

2.3 Pagel’s discrete dynamic method

3 Dynamic phylogenetic models

3.1 Formal model

3.2 Implementation

3.3 Synthetic example

3.4 Empirical applications

3.4.1 Brain size evolution in primates

3.4.2 Evolution of social complexity across human societies

4 Conclusion

References

- Bielby, Jon, Geogina M Mace, Olaf RP Bininda-Emonds, Marcel Cardillo, John L Gittleman, Katherine E Jones, C David L Orme, and A Purvis. 2007. “The Fast-Slow Continuum in Mammalian Life History: An Empirical Reevaluation.” *The American Naturalist* 169 (6): 748–57. <https://doi.org/10.1086/516847>.

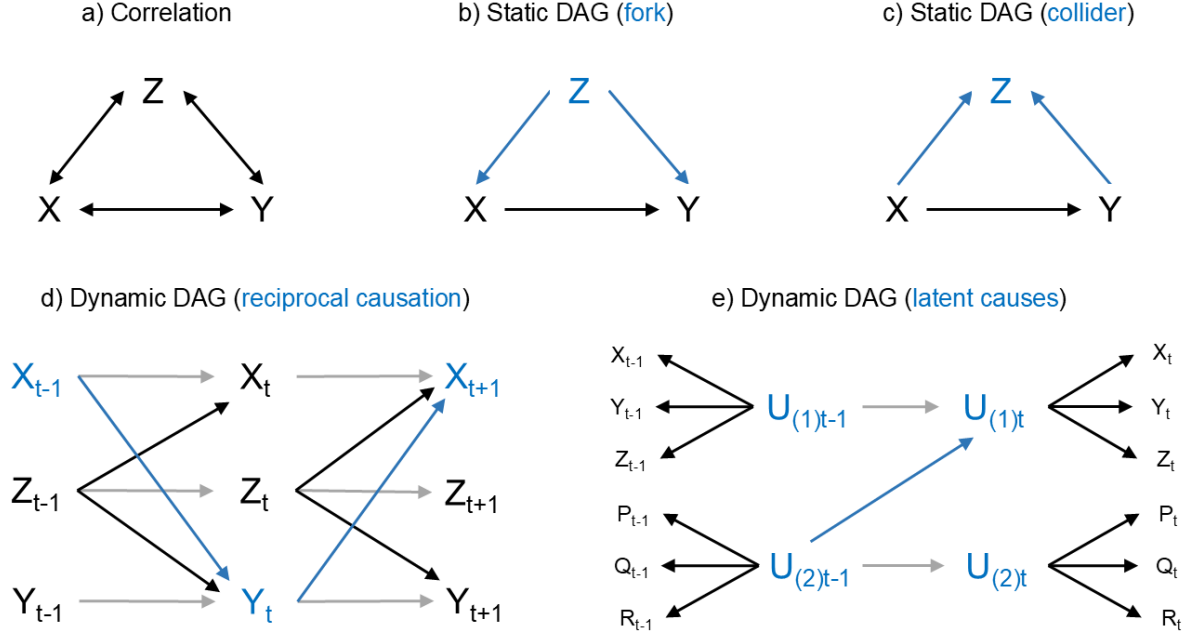


Figure 1: Explanatory models of trait coevolution.

Note: Examples of distinct formal approaches to describing and explaining patterns of coevolution among traits (bold letters), with important properties represented in each graph highlighted by blue arrows and text. Approaches range in complexity from (a) simple models of phylogenetic correlations (indicated by bidirectional arrows), which are useful for predictive purposes but fail to explain the causes of trait associations; to (b-e) directed acyclic graph models, which can be used to directly test the causal effects (directed arrows) driving trait associations across a phylogeny. As explained in Box 2 below, explicit causal models are crucial for deciding which traits should be included or excluded from a multivariate analysis to avoid potential biases due to phenomena such as so-called forks (b) and colliders (c). Causal models can also be further distinguished by whether they model relationships among traits as static (b-c) or dynamic (d-e) effects. Only dynamic models can be used to account for feedback processes ($t_{-1} \rightarrow t \rightarrow t_{+1}$) generated by reciprocal causation (blue arrows) among traits and autoregressive effects within traits (grey arrows) over time (d). For high dimensional problems, it is important to also consider whether inclusion of latent causes into the dynamic DAG (e), capturing dimensions of evolutionary integration among multiple traits, can provide a more parsimonious and theoretically insightful causal model (see [subsection 3.4](#)).

- Carpenter, Bob, Andrew Gelman, Matthew D Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus A Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. 2017. "Stan: A Probabilistic Programming Language." *Journal of Statistical Software* 76. <https://doi.org/10.18637/jss.v076.i01>.
- Clayton, Dale H, and Peter Cotgreave. 1994. "Relationship of Bill Morphology to Grooming Behaviour in Birds." *Animal Behaviour* 47 (1): 195–201. <https://doi.org/10.1006/anbe.1994.1022>.
- DeCasien, Alex R, Scott A Williams, and James P Higham. 2017. "Primate Brain Size Is Predicted by Diet but Not Sociality." *Nature Ecology & Evolution* 1 (5): 0112. <https://doi.org/10.1038/s41559-017-0112>.
- Dunn, Casey W, and Joseph F Ryan. 2015. "The Evolution of Animal Genomes." *Current Opinion in Genetics & Development* 35: 25–32. <https://doi.org/10.1016/j.gde.2015.08.006>.
- Fitzpatrick, John L, Robert Montgomerie, Julie K Desjardins, Kelly A Stiver, Niclas Kolm, and Sigal Balshine. 2009. "Female Promiscuity Promotes the Evolution of Faster Sperm in Cichlid Fishes." *Proceedings of the National Academy of Sciences* 106 (4): 1128–32. <https://doi.org/10.1073/pnas.080999010>.
- Garamszegi, László Zsolt. 2014. *Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology: Concepts and Practice*. Springer.
- Garland, Jr, Theodore, Albert F. Bennett, and Enrico L. Rezende. 2005. "Phylogenetic Approaches in Comparative Physiology." *Journal of Experimental Biology* 208 (16): 3015–35. <https://doi.org/10.1242/jeb.01745>.
- Gonzalez-Voyer, Alejandro, and Achaz von Hardenberg. 2014. "An Introduction to Phylogenetic Path Analysis." In *Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology: Concepts and Practice*, edited by László Zsolt Garamszegi, 201–29. Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-43550-2_8.
- Grafen, Alan. 1989. "The Phylogenetic Regression." *Philosophical Transactions of the Royal Society of London. B, Biological Sciences* 326 (1233): 119–57. <https://doi.org/10.1098/rstb.1989.0106>.
- Hadfield, JD, and S Nakagawa. 2010. "General Quantitative Genetic Methods for Comparative Biology: Phylogenies, Taxonomies and Multi-Trait Models for Continuous and Categorical Characters." *Journal of Evolutionary Biology* 23 (3): 494–508. <https://doi.org/10.1111/j.1420-9101.2009.01915.x>.
- Hardenberg, Achaz von, and Alejandro Gonzalez-Voyer. 2013. "Disentangling Evolutionary Cause-Effect Relationships with Phylogenetic Confirmatory Path Analysis." *Evolution* 67 (2): 378–87. <https://doi.org/10.1111/j.1558-5646.2012.01790.x>.
- Harvey, Paul H, and Mark D Pagel. 1991. *The Comparative Method in Evolutionary Biology*. Oxford University Press.
- Mace, Ruth, and Clare J Holden. 2005. "A Phylogenetic Approach to Cultural Evolution." *Trends in Ecology & Evolution* 20 (3): 116–21.
- MacLean, Evan L, Luke J Matthews, Brian A Hare, Charles L Nunn, Rindy C Anderson, Filippo Aureli, Elizabeth M Brannon, et al. 2012. "How Does Cognition Evolve? Phylogenetic Comparative Psychology." *Animal Cognition* 15: 223–38. <https://doi.org/10.1007/s10071-011-0448-8>.
- Navalón, Guillermo, Jen A Bright, Jesús Marugán-Lobón, and Emily J Rayfield. 2019. "The Evolutionary Relationship Among Beak Shape, Mechanical Advantage, and Feeding Ecology in Modern Birds." *Evolution* 73 (3): 422–35. <https://doi.org/10.1111/evo.13655>.
- Navarrete, Ana F, Simon M Reader, Sally E Street, Andrew Whalen, and Kevin N Laland. 2016. "The Coevolution of Innovation and Technical Intelligence in Primates." *Philosophical Transactions of the Royal Society B: Biological Sciences* 371 (1690): 20150186. <https://doi.org/10.1098/rstb.2015.0186>.
- Nunn, Charles L. 2011. *The Comparative Approach in Evolutionary Anthropology and Biology*. University of Chicago Press.
- O'Connor, Emily A, and Charlie K Cornwallis. 2022. "Immunity and Lifespan: Answering Long-Standing Questions with Comparative Genomics." *Trends in Genetics* 38 (7): 650–61. <https://doi.org/10.1016/j.tig.2022.02.014>.
- Pagel, Mark. 1994. "Detecting Correlated Evolution on Phylogenies: A General Method for the Comparative Analysis of Discrete Characters." *Proceedings of the Royal Society of London. Series B: Biological Sciences* 255 (1342): 37–45. <https://doi.org/10.1098/rspb.1994.0006>.
- Ringen, Erik, Jordan Scott Martin, and Adrian Jaeggi. 2021. "Novel Phylogenetic Methods Reveal That Resource-Use Intensification Drives the Evolution of "Complex" Societies." *EcoEvoRxiv*. <https://doi.org/10.32942/osf.io/wfp95>.
- Salguero-Gómez, Roberto, Owen R Jones, Eelke Jongejans, Simon P Blomberg, David J Hodgson, Cyril

- Mbeau-Ache, Pieter A Zuidema, Hans De Kroon, and Yvonne M Buckley. 2016. “Fast–Slow Continuum and Reproductive Strategies Structure Plant Life-History Variation Worldwide.” *Proceedings of the National Academy of Sciences* 113 (1): 230–35. <https://doi.org/10.1073/pnas.1506215112>.
- Sheehan, Oliver, Joseph Watts, Russell D Gray, Joseph Bulbulia, Scott Claessens, Erik J Ringen, and Quentin D Atkinson. 2023. “Coevolution of Religious and Political Authority in Austronesian Societies.” *Nature Human Behaviour* 7 (1): 38–45. <https://doi.org/10.1038/s41562-022-01471-y>.
- Shipley, Bill. 2016. *Cause and Correlation in Biology: A User’s Guide to Path Analysis, Structural Equations and Causal Inference with R*. Cambridge university press.
- Symonds, Matthew R. E., and Simon P. Blomberg. 2014. “A Primer on Phylogenetic Generalised Least Squares.” In *Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology: Concepts and Practice*, edited by László Zsolt Garamszegi, 105–30. Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-43550-2_5.
- Thayer, Zaneta M, Meredith A Wilson, Andrew W Kim, and Adrian V Jaeggi. 2018. “Impact of Prenatal Stress on Offspring Glucocorticoid Levels: A Phylogenetic Meta-Analysis Across 14 Vertebrate Species.” *Scientific Reports* 8 (1): 1–9. <https://doi.org/10.1038/s41598-018-23169-w>.
- Watts, Joseph, Oliver Sheehan, Quentin D Atkinson, Joseph Bulbulia, and Russell D Gray. 2016. “Ritual Human Sacrifice Promoted and Sustained the Evolution of Stratified Societies.” *Nature* 532 (7598): 228–31. <https://doi.org/10.1038/nature17159>.