University of Reading

Phylogenetic non-independence in rates of trait evolution

Article

Accepted Version

Sakamoto, M. and Venditti, C. ORCID: https://orcid.org/0000-0002-6776-2355 (2018) Phylogenetic non-independence in rates of trait evolution. Biology Letters, 14 (10). 20180502. ISSN 1744-957X doi: 10.1098/rsbl.2018.0502 Available at https://centaur.reading.ac.uk/79674/

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.

To link to this article DOI: http://dx.doi.org/10.1098/rsbl.2018.0502

Publisher: The Royal Society

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the End User Agreement.

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

University of Reading

Reading's research outputs online

Phylogenetic non-independence in rates of trait evolution

RUNNING HEAD: Phylogenetic non-independence in rates

Manabu Sakamoto[1*] and Chris Venditti[1*]

[1]School of Biological Sciences, University of Reading, Reading, UK, RG6 6BX

[*]CORRESPONDING AUTHOR CONTACT DETAILS:

m.sakamoto@reading.ac.uk

c.d.venditti@reading.ac.uk

ABSTRACT

Statistical non-independence of species' biological traits is recognized in most traits under selection. Yet, whether or not the evolutionary rates of such biological traits are statistically non-independent remains to be tested. Here we test the hypothesis that phenotypic evolutionary rates are non-independent, i.e. contain phylogenetic signal, using empirical rates of evolution in three separate traits: body mass in mammals; beak shape in birds; and bite force in amniotes. Specifically, we test whether rates are non-independent throughout the evolutionary history of each tree. We find evidence for phylogenetic signal in evolutionary rates in all three case studies. While phylogenetic signal diminishes deeper in time, this is reflective of statistical power owing to small sample and effect sizes. When effect size is large, e.g., owing to the presence of fossil tips, we detect high phylogenetic signals even in deeper time slices. Thus, we recommend that rates be treated as being non-independent throughout the evolutionary history of the group of organisms under study, and any summaries or analyses of rates through time – including associations of rates with traits – need account for the undesired effects of shared ancestry.

KEY WORDS: evolutionary rates; trait evolution; phylogeny; phylogenetic comparative methods; phylogenetic signal

Descent with modification [1] is of fundamental importance to evolution and is recognized in most traits under selection. Through evolutionary time, trait values will be more similar in closely related species compared to distantly related species, since the variance of trait values will be proportional to the divergence in evolutionary time [2]. This equates to shared ancestry, i.e. phylogeny. The degree to which shared ancestry affects biological traits can thus be described by the proportion of variance in trait data across a comparative sample of species that can be explained by phylogenetic relations, or phylogenetic signal – e.g., *K* [3] or λ [4]. This has statistical implications, i.e., phylogenetic non-independence.

While acknowledgement of phylogenetic non-independence in phenotypic trait data has become common in comparative studies [5], it is not so for the rates of evolution (how fast organisms' characteristics evolve). As rates are often used as proxies for adaptations [6, 7], it is of immense importance that we understand their statistical properties, in particular, phylogenetic non-independence. However, we have not been able to identify any study in the literature that explicitly tests for phylogenetic signal in phenotypic evolutionary rates aside from rare instances in which this was implied [8].

Here, we test whether evolutionary rates contain phylogenetic signal using three empirical case studies: body mass in mammals [9]; beak shape in birds [10]; and bite force in terrestrial amniotes (ESM). Our basic premise is that if phylogenetic signal is detected in rates, then rates evolve along the branches of a phylogenetic tree in proportion to the passage of time and that closely related species are more similar in rates than distantly related species. Naturally, this necessitates a non-homogenous distribution of rates across the branches of the tree – i.e., variable-rates of phenotypic trait evolution [8, 11-14].

MATERIALS AND METHODS

We obtained 100 sets of phenotypic evolutionary rates and the associated time-calibrated phylogenetic trees (time-trees) from the authors of the three case studies (ESM). As we would expect rates along all branches of a phylogenetic tree to be affected by shared ancestry, not just the tips, we tested for phylogenetic signal in rates along both terminal and internal branches, by time-slicing the phylogenetic tree. We time-sliced the three time-trees at 1-Myr intervals for the mammals and birds (167 and 109 time slices respectively), and at 5-Myr intervals for amniotes (65 time slices) (see ESM for details). The latter interval was chosen for amniotes owing to their much longer evolutionary history (approx. 350 Myr) compared to mammals and birds. For each time-sliced tree, we matched the terminal branches to the corresponding branches in the complete time-tree (Fig S1). We then assigned the corresponding rates to those terminal branches on the time-sliced tree as tip trait values. We fitted a maximum likelihood (ML) phylogenetic generalized least squares (GLS) model in BayesTraits V3 to estimate phylogenetic signal λ in rates at the tips for each time-slice (GLS_{λ}). We tested GLS_{λ} against the null model in which λ is fixed to 0 ($GLS_{\lambda=0}$) as the likelihood ratio (LR_{λ}) between GLS_{λ} and $GLS_{\lambda=0}$ and determined significance using the χ^2 distribution (df=1). When λ was significant in >95% of the sample in any given time slice, we determined that phylogenetic signal was present in that time slice. We also compared the significance of an alternative model in which λ is fixed to 1 ($GLS_{\lambda=1}$). The root estimate α of a $GLS_{\lambda=0}$ model is the equivalent of estimating the non-phylogenetic mean rate, while GLS_{λ} and $GLS_{\lambda=1}$ estimate the phylogenetically corrected mean rates.

RESULTS

Overall, phylogenetic signal at the tips of the complete time trees are high (body mass in mammals, median $\lambda = 0.926$; beaks in birds, median $\lambda = 0.729$; and bite force in amniotes, $\lambda = 1$), providing evidence for strong effects of shared ancestry in rates of phenotypic trait evolution along the terminal branches. Phylogenetic signal in rates are generally high and significant in at least 95% of the sample in younger time slices – younger than: 48 Myr ago (mammals); 45 Myr ago (birds); and 30 Myr ago (amniotes) (Fig. 1; ESM). Phylogenetic signal depreciates (drops in strength and significance) rapidly in deeper time slices (Fig. 1; ESM). Fixing λ to 1 ($GLS_{\lambda=1}$) result in qualitatively similar patterns across time slices compared to when λ is estimated (GLS_{λ}) (Fig. 1), but depreciation of λ start at younger time slices compared to GLS_{λ} (Fig. 1).

DISCUSSION

Through our time-sliced GLS models on three datasets, we demonstrate that evolutionary rates of phenotypic traits are indeed phylogenetically non-independent – λ is significant and high, both along the terminal and internal branches (Fig. 1). Crucially, although λ ceases to be significant in deeper time slices in all trees tested (Fig. 1), this reduction in phylogenetic signal most likely depends on two aspects of the rates in the focal time slice: 1) number of tips [15] and 2) rate heterogeneity (ESM). Both reflect issues of statistical power with the former concerning sample size (as determined through simulations; ESM) and the latter effect size (as

100 evident from the effects of fossil tips; Fig 1C; ESM). Incidentally, un-sampled tips of any sort
101 (not just fossils) will likely increase rate heterogeneity should they be sampled. Additionally,
102 information contained at the tips of an ultrametric tree (e.g., trait values) is expected to be lost
103 progressively deeper in the tree (proportional to the phylogenetic variance-covariance
104 structure) as subsequent evolution towards the tips overprints ancestral information – this is an
105 issue plaguing phylogenetic comparative methods in general. Furthermore, since rates are
106 estimated from the phylogeny using models with constant rate evolution as the underlying
107 process of evolution, the resulting rates would inevitably contain phylogenetic signal. Whether
108 this is true or not, this does not alter (rather it reinforces) our argument that inferred rates
109 contain phylogenetic signal (regardless of the reason) and crucially that all downstream
110 summaries and analyses of rates thus must account for phylogenetic non-independence. Thus,
111 we argue that it is safest to assume that phylogenetic signal will be present and strong in
112 deeper time slices [8].

113 An important implication here is that as rates will be statistically non-independent at
114 various time intervals throughout the history of the clade of interest, patterns gleaned from
115 simple summaries (e.g., interval means) of rates-through-time (RTT) can potentially be
116 misleading. Simple RTT plots are prevalent in recent literature [e.g., 10, 16, 17-21], the profiles
117 of which routinely interpreted at face value, with peaks and troughs representing periods of
118 bursts and declines in rates [16, 17, 19]. However, accounting for phylogenetic non-
119 independence by assuming strong phylogenetic signal uniformly across all time slices [8] – i.e.
120 phylogenetic mean α from our $GLS_{\lambda=1}$ models across time slices – results in phylo-RTT profiles
121 that are often different from those of non-phylogenetic RTT (Fig. 2). Thus, non-phylogenetic

RTT profiles cannot be taken at face value without knowledge of phylogenetic signal through time. More crucially, this implies that statistical analyses of rates need also account for phylogenetic non-independence. Testing hypotheses of external influences (ecological or environmental) on rates of evolution would require the application of appropriate phylogenetic statistical methods – e.g. phylogenetic regression models [14, 22]. Not doing so will run the risk of resulting in misleading statistical results [2, 5].

As phenotypic evolutionary rates have been interpreted as reflecting the intensity of natural selection [6, 14], that they contain phylogenetic signal implies that ancestors and descendants as well as closely related species either: 1) share intrinsic mechanisms for selection responsiveness (e.g., genetic predisposition); 2) share similar levels of extrinsic selection pressures (e.g., similar ecological niches, environments, etc); or 3) both. Two (or more) species descended from a parent species would be expected to start their respective independent evolution with the same level of intrinsic responsiveness as well as extrinsic selection pressures, and thus at the phenotypic evolutionary rate, of the parent species. The daughter species then would be subject to independent genetic mutations and selection pressures depending on their respective environments.

However, this is not to say that descendent rates are rigidly constrained by ancestry; exceptional rate shifts along individual branches are widely observed in many traits across various groups of organisms [8, 9, 11, 14, 23]. Such exceptional rate shifts can often be orders of magnitude greater than the background rate and occur instantaneously (with respect to geological time) such that the effects of ancestry may be marginal.

In conclusion, our analyses demonstrate that rates of phenotypic evolution estimated from phylogenetic trees using models of trait evolution are statistically non-independent (most likely owing to shared ancestry), across the tips and through time – we posit that our results are conservative with phylogenetic signal actually being more prevalent. Thus, we recommend that phylogenetic non-independence be accounted for in summaries and analyses of evolutionary rates through time, using appropriate phylogenetic comparative methods.

LITERATURE CITED

[1] Darwin, C. 1859 *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. First Edition ed. London, UK.

[2] Felsenstein, J. 1985 Phylogenies and the Comparative Method. *American Naturalist* **125**, 1-15. (doi:Doi 10.1086/284325).

[3] Blomberg, S.P., Garland, T. & Ives, A.R. 2003 Testing for phylogenetic signal in comparative data: Behavioral traits are more labile. *Evolution* **57**, 717--745.

[4] Pagel, M. 1997 Inferring evolutionary processes from phylogenies. *Zoologica Scripta* **26**, 331-348. (doi:10.1111/j.1463-6409.1997.tb00423.x).

[5] Harvey, P.H. & Pagel, M.D. 1991 *The comparative method in evolutionary biology*, Oxford University Press.

[6] Simpson, G.G. 1944 *Tempo and mode in evolution*. New York, Columbia University Press; 237 p.

[7] Pagel, M. 1994 The adaptationist wager. In *Phylogenetics and Ecology* (eds. P. Eggleton & R.I. Vane-Wright), pp. 29-51. London, Academic Press.

[8] Venditti, C., Meade, A. & Pagel, M. 2011 Multiple routes to mammalian diversity. *Nature* **479**, 393-396. (doi:10.1038/nature10516).

[9] Baker, J., Meade, A., Pagel, M. & Venditti, C. 2015 Adaptive evolution toward larger size in mammals. *Proceedings of the National Academy of Sciences* **112**, 5093-5098. (doi:10.1073/pnas.1419823112).

[10] Cooney, C.R., Bright, J.A., Capp, E.J.R., Chira, A.M., Hughes, E.C., Moody, C.J.A., Nouri, L.O., Varley, Z.K. & Thomas, G.H. 2017 Mega-evolutionary dynamics of the adaptive radiation of birds. *Nature* **542**, 344-347. (doi:10.1038/nature21074 http://www.nature.com/nature/journal/v542/n7641/abs/nature21074.html#supplementary-information).

[11] Eastman, J.M., Alfaro, M.E., Joyce, P., Hipp, A.L. & Harmon, L.J. 2011 A Novel Comparative Method for Identifying Shifts in the Rate of Character Evolution on Trees. *Evolution* **65**, 3578-3589. (doi:10.1111/j.1558-5646.2011.01401.x).

[12] Rabosky, D.L. 2014 Automatic Detection of Key Innovations, Rate Shifts, and Diversity-Dependence on Phylogenetic Trees. *Plos One* **9**, e89543. (doi:10.1371/journal.pone.0089543).

[13] Elliot, M.G. & Mooers, A.Ø. 2014 Inferring ancestral states without assuming neutrality or gradualism using a stable model of continuous character evolution. *Bmc Evol Biol* **14**, 226. (doi:10.1186/s12862-014-0226-8).

[14] Baker, J., Meade, A., Pagel, M. & Venditti, C. 2016 Positive phenotypic selection inferred from phylogenies. *Biological Journal of the Linnean Society* **118**, 95-115. (doi:10.1111/bij.12649).

[15] Freckleton, R.P., Harvey, P.H. & Pagel, M. 2002 Phylogenetic analysis and comparative data: A test and review of evidence. *American Naturalist* **160**, 712--726.

[16] Close, Roger A., Friedman, M., Lloyd, Graeme T. & Benson, Roger B.J. 2015 Evidence for a Mid-Jurassic Adaptive Radiation in Mammals. *Current Biology*. (doi:10.1016/j.cub.2015.06.047).

[17] Felice, R.N. & Goswami, A. 2018 Developmental origins of mosaic evolution in the avian cranium. *Proc Natl Acad Sci U S A* **115**, 555-560. (doi:10.1073/pnas.1716437115).

[18] Lloyd, G.T., Wang, S.C. & Brusatte, S.L. 2012 Identifying heterogeneity in rates of morphological evolution: Discrete character change in the evolution of lungfish (Sarcopterygii; Dipnoi). *Evolution* **66**, 330-348. (doi:10.1111/j.1558-5646.2011.01460.x).
[19] Hopkins, M.J. & Smith, A.B. 2015 Dynamic evolutionary change in post-Paleozoic echinoids and the importance of scale when interpreting changes in rates of evolution. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 3758-3763. (doi:10.1073/pnas.1418153112).
[20] Rabosky, D.L., Donnellan, S.C., Grundler, M. & Lovette, I.J. 2014 Analysis and Visualization of Complex Macroevolutionary Dynamics: An Example from Australian Scincid Lizards. *Systematic Biology* **63**, 610-627. (doi:10.1093/sysbio/syu025).
[21] Evans, A.R., Jones, D., Boyer, A.G., Brown, J.H., Costa, D.P., Ernest, S.K.M., Fitzgerald, E.M.G., Fortelius, M., Gittleman, J.L., Hamilton, M.J., et al. 2012 The maximum rate of mammal evolution. *Proceedings of the National Academy of Sciences* **109**, 4187-4190.
[22] Orme, D., Freckleton, R., Thomas, G., Petzoldt, T., Fritz, S., Isaac, N. & Pearse, W. 2013 caper: Comparative Analyses of Phylogenetics and Evolution in R. R package version 0.5.2. (R package version 0.5.2 ed.
[23] Chira, A.M. & Thomas, G.H. 2016 The impact of rate heterogeneity on inference of phylogenetic models of trait evolution. *Journal of Evolutionary Biology* **29**, 2502-2518. (doi:10.1111/jeb.12979).
[24] Sakamoto, M. & Venditti, C. 2018 Phylogenetic non-independence in rates of trait evolution. (osf.io/pn4ma.

ACKNOWLEDGEMENTS:

We thank Joanna Baker, Ciara O'Donovan, Renske Gudde, Andrew Meade and Mark Pagel for invaluable discussion and technical support, which helped improve the research and manuscript.

FUNDING:

This work was supported by the Leverhulme Trust (RPG-2013-185 and RPG-2017-071).

AUTHOR CONTRIBUTIONS:

All authors made substantial contributions to conception and design, acquisition of data, analysis and interpretation of data; drafted and revised the article critically for important intellectual content; made final approval of the version to be published; and made agreements to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

DATA ACCESSIBILITY:

Supporting data are made available through the Open Science Framework (https://osf.io/pn4ma/?view_only=2de49c7ad61944ed97c373783a7d1956)[24], and described in ESM.

COMPETING INTERESTS:

The authors declare no competing interests.

FIGURE LEGENDS

Figure 1. Phylogenetic signal in rates of trait evolution through time.

Phylogenetic signal (λ) was estimated across time sliced phylogenetic trees (top row) in three independent data sets: A, mammalian body mass; B, avian beak shape; and C, aminote bite force. Faint lines represent each of the 100 samples with the bold line representing the median λ. The percentage of the sample in which LR_{λ} (likelihood ratio between GLS_{λ} and $GLS_{\lambda=0}$) was significant is shown for each time slice (second row). Further, the fit of $GLS_{\lambda=1}$ is shown as the percentage of the sample in which $LR_{\lambda=1}$ (likelihood ratio between $GLS_{\lambda=1}$ and $GLS_{\lambda=0}$) was significant for each time slice (third row). Red dashed line represents the 95% threshold. Blue dashed line (top) represents the time slice for the 95% threshold as determined through simulations (Fig. S3). The relationship between the percentage of significant λ and N_{Tips} (bottom row) shows a clear drop off in the percentage from 95% of the sample (red box).

Figure 2. Mean evolutionary rates through time compared to phylogenetically corrected mean rates. Simple mean values of evolutionary rates at each time slice across the three datasets (A, mammalian body mass; B, avian beak shape; C, amniote bite force) show distinctive patterns of rates through time. However, these patterns are far less prominent in phylogenetically corrected mean rates (α) through time. α are the phylogenetic root estimates of the $GLS_{\lambda=1}$ model. Faint lines represent each MCMC run while the bold line shows the median value for each time slice.