

# Incompletely resolved phylogenetic trees inflate estimates of phylogenetic conservatism

T. JONATHAN DAVIES,<sup>1,6</sup> NATHAN J. B. KRAFT,<sup>2</sup> NICOLAS SALAMIN,<sup>3,4</sup> AND ELIZABETH M. WOLKOVICH<sup>5</sup>

<sup>1</sup>Department of Biology, McGill University, 1205 ave Docteur Penfield, Montreal, Quebec H3A 1B1 Canada

<sup>2</sup>Biodiversity Research Centre, University of British Columbia, Vancouver V6T 1Z4 Canada

<sup>3</sup>Department of Ecology and Evolution, University of Lausanne, 1015 Lausanne, Switzerland

<sup>4</sup>Swiss Institute of Bioinformatics, Quartier Sorge, 1015 Lausanne, Switzerland

<sup>5</sup>Division of Biological Sciences, University of California–San Diego, La Jolla, California 92093 USA

**Abstract.** The tendency for more closely related species to share similar traits and ecological strategies can be explained by their longer shared evolutionary histories and represents phylogenetic conservatism. How strongly species traits co-vary with phylogeny can significantly impact how we analyze cross-species data and can influence our interpretation of assembly rules in the rapidly expanding field of community phylogenetics. Phylogenetic conservatism is typically quantified by analyzing the distribution of species values on the phylogenetic tree that connects them. Many phylogenetic approaches, however, assume a completely sampled phylogeny: while we have good estimates of deeper phylogenetic relationships for many species-rich groups, such as birds and flowering plants, we often lack information on more recent interspecific relationships (i.e., within a genus). A common solution has been to represent these relationships as polytomies on trees using taxonomy as a guide. Here we show that such trees can dramatically inflate estimates of phylogenetic conservatism quantified using S. P. Blomberg et al.'s *K* statistic. Using simulations, we show that even randomly generated traits can appear to be phylogenetically conserved on poorly resolved trees. We provide a simple rarefaction-based solution that can reliably retrieve unbiased estimates of *K*, and we illustrate our method using data on first flowering times from Thoreau's woods (Concord, Massachusetts, USA).

**Key words:** Blomberg's *K*; ecophylogenetics; niche conservatism; phylogeny; *Phylomatic*; polytomies.

## INTRODUCTION

Closely related species frequently resemble each other in form and function, reflecting their shared evolutionary histories and the inheritance of traits from a common ancestor. This trend is sometimes referred to as phylogenetic conservatism (Harvey and Pagel 1991), because the expected covariation among species can be directly estimated from their phylogeny. Ecologists are increasingly making use of the information contained within phylogenetic trees (Ackerly 2009), and phylogenetic conservatism is an important foundation for many research areas in ecology (Wiens et al. 2010). Although phylogenetic conservatism has proven a somewhat slippery concept to define analytically (see, e.g., Losos 2008, Wiens 2008), for purposes here, we equate phylogenetic conservatism with phylogenetic signal, the tendency for closely related species to be more similar than expected by chance, and we use the two terms interchangeably. In the emerging field of ecophylogenetics, the quantification of phylogenetic signal is a critical component for inferring community assembly

rules. Importantly, the ecological processes implicated in the assembly of a given community can change depending on the strength of conservatism of key traits. (Webb et al. 2002, Cavender-Bares et al. 2009, Vamosi et al. 2009).

There exist many statistical approaches for evaluating phylogenetic conservatism (see Revell et al. 2008). For continuous traits, species are typically assumed to diverge over time in a manner analogous to a random walk, assuming a Brownian motion model of evolution, with variance increasing in proportion to the square root of the sum of the evolutionary distance separating taxa (Felsenstein 1985). A commonly used metric of phylogenetic conservatism is Blomberg et al.'s *K* statistic (Blomberg et al. 2003), hereon Blomberg's *K*. Blomberg's *K* reflects the observed phenotypic similarity in taxa relative to expectations from a Brownian model given the phylogenetic topology that connects them. Importantly, Blomberg's *K* is not only comparable across traits but also across phylogenetic trees, so that strength of phylogenetic signal can be compared directly among different clades (Blomberg et al. 2003). Simulations demonstrate that Blomberg's *K* performs well on completely sampled phylogenetic trees (Blomberg et al. 2003); however, to date, its performance has not been

Manuscript received 28 July 2011; revised 19 October 2011; accepted 24 October 2011. Corresponding Editor: N. J. Sanders.

<sup>6</sup> E-mail: j.davies@mcgill.ca

evaluated on the type of phylogenetic data most commonly available to ecologists.

Recent, rapid advances in molecular sequencing technology have provided a wealth of phylogenetic data that has allowed ecologists to address questions on a large scale, using phylogenetic trees including dozens to hundreds of species. For many species-rich clades, such as flowering plants and birds, we have a good understanding of deeper phylogenetic relationships (e.g., see Angiosperm Phylogeny Group [1998, 2003, 2009] and the DNA-DNA hybridization studies of birds through the 1980s and early 1990s by Charles Sibley and Jon Edward Ahlquist [Sibley and Ahlquist 1983, 1987, 1990]). However, phylogenetic relationships between the many thousands of species within higher taxonomic groups remain largely unresolved (Hodkinson and Parnell 2007). One common approach to placing missing species within a phylogenetic framework is to use taxonomy as a guide. Species are added to a backbone phylogeny, typically as a basal polytomy subtending from the minimally inclusive node representing the clade/taxon in which they are known to be members (Fig. 1). As a tool for ecologists, this approach has been automated in the Phylomatic online software (Webb and Donoghue 2005). Adopted most rapidly by community ecologists interested in exploring phylogenetic structure in species assemblages, Phylomatic-derived phylogenetic trees have become common in the ecological literature. Researchers can optionally resolve nodes to reflect new information not present in the backbone phylogeny. Nonetheless, for species-rich data sets, some information on phylogenetic relationships is almost always lacking, and such trees tend to have a particular structure, reflecting the method of their assemblage, with a relatively well-resolved backbone and many terminal polytomies.

Recently, there has been growing evidence for apparent incongruence in phylogenetic metrics derived from studies using phylogenetic trees reconstructed using different methods. In an analysis of phylogenetic community structure of neotropical forest trees, Kress et al. (2009) suggested a bias toward inflated type II errors for Phylomatic trees. Here, we suggest estimates of phylogenetic signal might also be biased. In one example, Uriarte and colleagues (Uriarte et al. 2010) reported little or no phylogenetic signal for tropical forest traits using Blomberg's  $K$ , contrasting with a previous analysis of the same community, but using a slightly different metric to capture signal (Swenson et al. 2007). Uriarte and colleagues used a DNA barcode phylogeny that was largely resolved, while the analysis by Swenson et al. (2007) relied on a Phylomatic phylogeny. Similarly, in a study of plant invasion, Cadotte et al. (2009) reported equally disparate estimates of phylogenetic conservatism between molecular vs. Phylomatic tree topologies, with  $K$  values for the latter being several orders of magnitude greater. Furthermore, in the complete absence of information

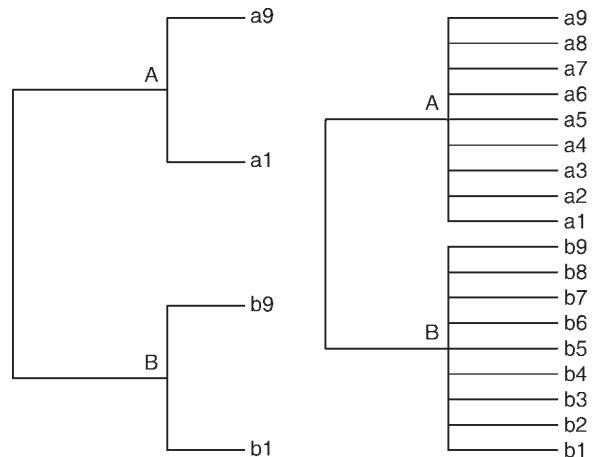


FIG. 1. Representation of a phylogenetic tree with two sister clades, A and B, corresponding to higher taxa, containing several species, indicated by lowercase tip labels. When phylogenetic relationships within higher taxa are unknown (left-hand tree), excluded species can be placed in the phylogeny as a basal polytomy subtending from the minimally inclusive node representing the taxon in which they are known to be members (right-hand tree).

on phylogenetic tree structure, where the tree is represented as an equal-length polytomy,  $K$  is constrained to equal 1 (Revell et al. 2008). These trends beg the question of whether there is a general bias toward overestimating phylogenetic conservatism in poorly resolved trees. Here we test for the effects of tree resolution on estimates of trait conservatism using Blomberg's  $K$  statistic.

#### Phylogenetic simulations

Using computer simulations we demonstrate a striking relationship between tree resolution and phylogenetic conservatism as indexed by  $K$ . Simulations were performed in the R software environment (2.13; R Development Core Team 2011) using the libraries ape (Paradis et al. 2004), picante (Kembel et al. 2010), and geiger (Harmon et al. 2008).

To explore the effect of poor phylogenetic resolution on estimates of trait conservatism, we generated a series of 1000 completely resolved phylogenetic trees of size  $n = 128$ . Tree topologies and branch lengths were generated by randomly clustering taxa and assuming coalescent branching times (function `rcoal` in the R library `ape`). For each tree topology, we then simulated the evolution of a continuous trait along its branches assuming a Brownian motion model of trait change (function `rTraitCont`) with rate parameter,  $\sigma = 0.1$ , and calculated Blomberg's  $K$  (function `Kcalc` in the R library `picante`) as our index of phylogenetic conservatism.

Blomberg's  $K$  is expressed as the ratio  $MSE_0/MSE$ , where  $MSE_0$  represents the mean squared error of the measured traits from the phylogenetically correct mean, and  $MSE$  is the mean squared error of the data calculated using the variance-covariance matrix derived

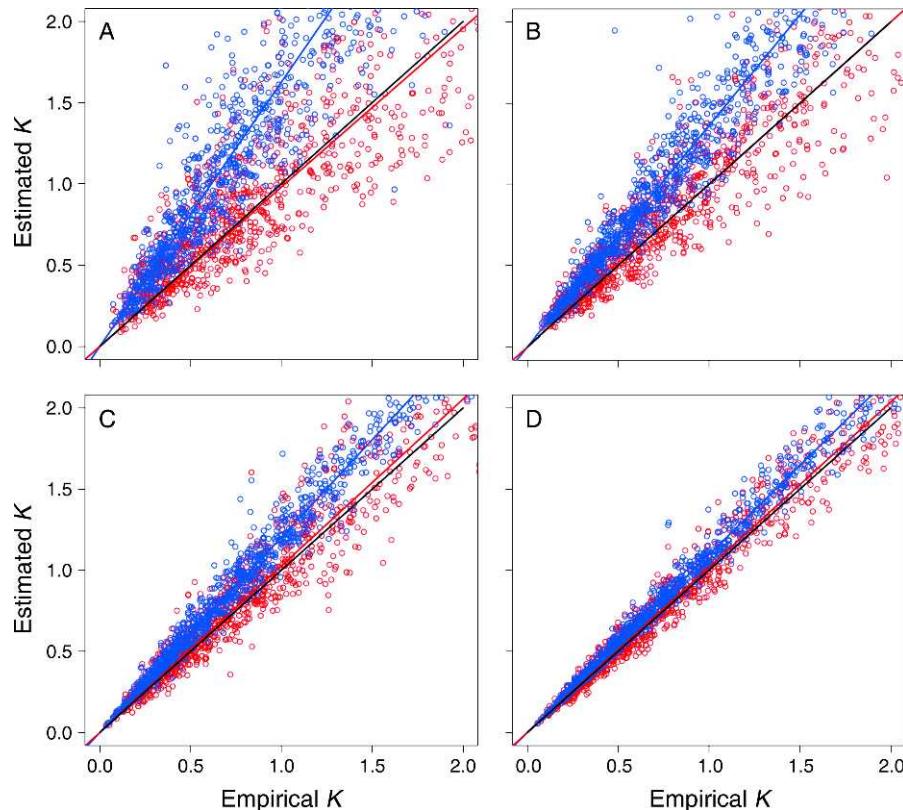


FIG. 2. Relationship between the metric of phylogenetic conservatism,  $K$ , from simulations on fully resolved trees vs.  $K$  from incompletely resolved (blue) and thinned (red) tree topologies. The black line indicates the 1:1 line where estimated and empirical  $K$  are identical. Phylogenetic resolution for the incompletely resolved trees: (A) 20%, (B) 40%, (C) 60%, and (D) 80% resolved nodes.

from the phylogeny (Blomberg et al. 2003). Large values of  $MSE_0/MSE$  indicate greater phylogenetic signal, but values estimated from different phylogenetic trees are not directly comparable. To allow comparisons across traits and trees, observed  $MSE_0/MSE$  is standardized by the expected ratio predicted under the assumption of Brownian motion on the given phylogeny, so that

$$K = \text{observed} \frac{MSE_0}{MSE} / \text{expected} \frac{MSE_0}{MSE}.$$

Values of  $K$  are bounded at 0, which indicates no phylogenetic structure in the measured traits, and converge on Brownian motion at 1. Values of  $K > 1$  imply that closely related taxa are more similar than expected from a model of Brownian motion evolution. By definition, the expected  $K$  value for the simulated traits is 1; however, because Brownian motion is a type of Markov model,  $K$  values for individual realizations will vary stochastically (mean  $K$  from simulations =  $1.02 \pm 0.89$  [mean  $\pm$  SD]). Next, we created random polytomies in each of the phylogenetic trees by collapsing nodes from the tips of the tree to the root, until only 40% of nodes were resolved. Our approach here was intended to replicate the branching structure typical of Phylomatic trees, with a high density of

terminal polytomies. Our index of phylogenetic conservatism was then estimated on the unresolved phylogenies using the same trait data.

To evaluate model sensitivity, we additionally explored alternate resolution thresholds (20%, 60%, and 80%), and a different set of starting trees generated by randomly splitting branches. For the latter set of tree topologies, branch lengths were assigned following Grafen (1989), where node height is proportional to the number of subtending leaves.

For all phylogenetic resolutions, the unresolved topologies led to significantly inflated estimates of phylogenetic conservatism, but with greater bias in the more unresolved topologies (slope  $\pm$  SE of regression of  $K$  from the fully resolved trees against  $K$  from the incompletely resolved trees =  $1.09 \pm 0.002$ ,  $1.20 \pm 0.004$ ,  $1.27 \pm 0.005$ , and  $1.63 \pm 0.014$ , for the coalescent trees at 80%, 60%, 40%, and 20% resolution, respectively; Fig. 2). Results were qualitatively similar for the random-split trees (Appendix: Fig. A1). In addition, even when the tip data were randomized, giving the expectation  $K \ll 1$  for no phylogenetic signal, estimates of phylogenetic conservatism converged on  $K = 1$  as nodes on the tree were sequentially collapsed (Fig. 3). However, the relationship between resolution and  $K$  was approximately logarithmic within the bounds  $1 > K > 0$ , such

that as resolution decreases,  $K$  increases exponentially (Fig. 3).

Finally, we tested for any bias in  $K$  with tree size by simulating a further set of 500 trees of size  $n = 1000$ , and then randomly pruned each tree to size  $n = 25, 50, 100, 125, 150, 200, 250, 300$ , and 500. Although variance in  $K$  was high among the various simulations (Appendix: Fig. A2), we observed no systematic bias in  $K$  with  $n$  (mean  $\pm$  SE for the slope of the regression of  $K$  against tree size =  $8.9 \times 10^{-7} \pm 1.11 \times 10^{-5}$ ).

#### *A simple solution for accurately estimating $K$ on poorly resolved phylogenies*

We propose a straightforward rarefaction procedure to accurately estimate phylogenetic conservatism on incompletely resolved trees. Our method randomly removes all but one taxon per terminal polytomy in the complete phylogeny, and estimates  $K$  on the smaller “thinned” tree. The procedure is repeated iteratively, and a distribution of  $K$  values is produced. Because  $K$  is independent from tree size (see *Phylogenetic simulations*), the thinned trees should produce unbiased estimates of  $K$ . We evaluated our method by applying it to our simulations (see *Phylogenetic simulations*; see Supplement for R scripts). For purposes here, we evaluated 10 thinned topologies per incompletely resolved tree. When true  $K$  is not known, as is the case for most empirical data, the number of iterations should be scaled appropriately with tree size and resolution so that mean and standard errors for  $K$  stabilize. Our simulations demonstrate that the true estimate of phylogenetic conservatism can be retrieved with a high degree of accuracy from the thinned trees (slopes of regressions of  $K$  from the fully resolved trees and mean  $K$  from the thinned trees = 0.98–1.03; Fig. 2), and that our method was robust to alternate tree topologies (Appendix: Fig. A2).

#### *Phylogenetic conservatism of flowering times in Thoreau's Woods*

In a recent paper, Willis et al. (2008) commented that “[t]he extent to which flowering-time response to temperature is shared among closely related species might have important consequences for community-wide patterns of species loss under rapid climate change.” Willis et al. (2008) use a robust randomization procedure to test for significant phylogenetic conservatism; however, the strength of conservatism was not reported. As in many recent studies, the phylogeny represented a composite tree constructed in Phylomatic, but additionally resolved to the level of genera within families. Species within genera were still represented as terminal polytomies. Here we use data from Miller-Rushing and Primack (2008) on mean first flowering dates in Thoreau's Woods (Concord, Massachusetts, USA) to directly quantify phylogenetic conservatism in this data set following Willis et al. (2008). First, we assume a minimally resolved phylogeny, simply accept-

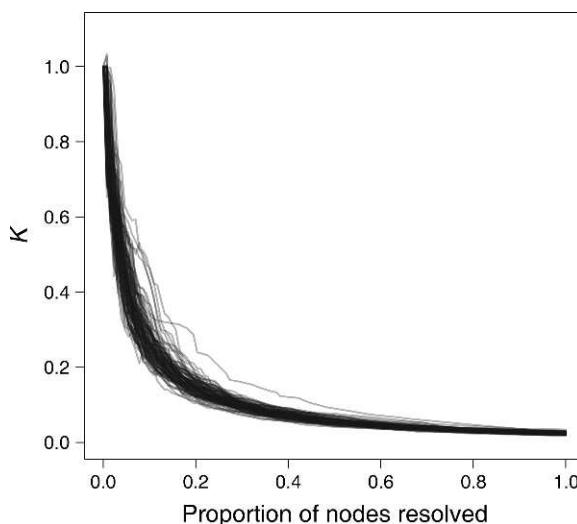


FIG. 3. Trend for increasing  $K$  with decreasing tree resolution simulated on 100 random tree topologies and randomly assigning trait values to tips. For completely unresolved trees,  $K$  is constrained to be 1 (Revell et al. 2008).

ing the backbone tree provided as the default option in Phylomatic. Second, we contrast results from the poorly resolved tree with those from the more resolved tree used by Willis and colleagues. For comparison, we next sequentially collapse nodes in the Willis tree, as described above, and reestimate  $K$  at each iteration. Finally, we use our subsampling method to determine our best estimate of  $K$ ; unfortunately it is not possible to compare our results to true  $K$  as we do not have a complete tree resolved to the species level.

The Phylomatic tree has approximately 45% of nodes resolved compared to a completely resolved tree. As predicted from our simulations, this tree returns the highest estimate of phylogenetic conservatism, with  $K = 0.48$ . Willis and colleagues performed an admirable job in resolving relationships, and their tree has 66% of nodes resolved. This increase in tree resolution results in a large decrease in our estimate of phylogenetic conservatism, with  $K = 0.28$ . By estimating  $K$  on the increasingly poorly resolved topologies, we show a monotonic relationship between  $K$  and tree resolution (Fig. 4). There is strong agreement in  $K$  estimated from the Phylomatic tree and for the Willis tree at equivalent resolution, even though the particular nodes subtending polytomies likely differ. However, we show that the additional effort invested in resolving polytomies in the Phylomatic tree was rewarded by returning an accurate estimate of  $K$ , with average  $K$  from the thinned trees (mean  $K = 0.29$ ) returning a similar value to that estimated directly from the Willis tree.

#### DISCUSSION

We show that poor phylogenetic resolution tends to inflate estimates of phylogenetic conservatism as quantified by Blomberg et al.'s  $K$  statistic. Our analysis is

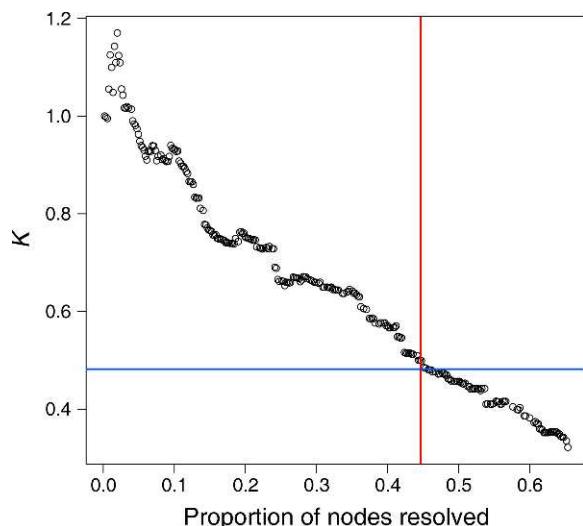


FIG. 4. Relationship between tree resolution (proportion of nodes resolved) and phylogenetic conservatism ( $K$ ) for mean first flowering dates in Thoreau's woods, estimated from the phylogenetic tree published by Willis et al. (2008). The vertical red line indicates the phylogenetic resolution of the default tree generated by Phylomatic (Webb and Donoghue 2005). The horizontal blue line indicates the value of  $K$  estimated directly from the Phylomatic tree.

focused upon a particular bias in estimates of phylogenetic signal. Nonetheless, this bias is important to recognize as strong conservatism is frequently implicit in the ecological and conservation literature today, where phylogenetic branch lengths are assumed to map closely to species functional and ecological diversities. In the rapidly expanding field of ecophylogenetics (recently reviewed by Emerson and Gillespie [2008], Vamوسي et al. [2009], and Cavender-Bares et al. [2009]), the interpretation of phylogenetic community structure depends critically on these estimates. In the conservation literature, metrics of phylogenetic diversity estimated from time-calibrated trees (Faith 1992, Crozier 1997) also make strong assumptions (frequently untested) on the phylogenetic conservatism of important traits. Erroneous estimates of phylogenetic conservatism might therefore mislead inferences drawn from such studies, a cause for concern given the common use of incompletely resolved phylogenetic trees evident in the literature.

Our simulations show that  $K$  values (for the same trait or for different traits) calculated on different phylogenies are not comparable if the degree of resolution differs between the trees. The ideal solution would be to invest in the resources necessary for producing fully resolved and well-supported phylogenies. However, for species-rich taxa, generating such data is logistically daunting, expensive and time consuming. Here, we have proposed a novel solution, using a rarefaction procedure. In our simulated data, we show that our method provides an unbiased estimate of true  $K$ , though we suggest that it is still far less preferable than properly

resolving the topologies. Our method does not evaluate, nor correct for, bias from polytomies nested deeper in the structure of the tree. For many species-rich clades, including angiosperms, higher-level phylogenetic relationships are relatively well resolved. Nonetheless, effort in resolving evolutionary relationships might be better focused at these deeper nodes, because thinning internal polytomies to a single descendent lineage might rapidly delete large segments of the phylogeny, at the extreme, thinning a polytomy at the root node would leave a single non-branching lineage.

Using our new method we quantified the strength of phylogenetic conservatism in first flowering times for the flora of Thoreau's woods. We show that phylogenetic conservatism is greatly inflated when estimated from a poorly resolved phylogeny, derived from the angiosperm family backbone tree, matching results from simulations. However, by using additional molecular analysis and information from the recent literature, Willis and colleagues (Willis et al. 2008) were able to resolve approximately two-thirds of nodes, which provided sufficient resolution to match our best estimate of  $K$  ( $K = 0.28\text{--}0.29$ ). It is some consolation that a completely resolved phylogeny is not necessary to approximate  $K$ , and from our simulations we suggest that there might be little bias in estimates for trees greater than 60% resolved. Nonetheless, many published phylogenies are less well resolved, and large trees generated by placing species within clades using taxonomies, as used in many analyses of community data, are likely to provide particularly poor estimates of  $K$ . We show that in such poorly resolved trees, even traits randomly assigned to species on the phylogeny will appear to demonstrate strong phylogenetic conservatism.

Ecological research, drawing on new phylogenetic data and methods, has advanced rapidly. It is now possible to address questions spanning evolutionary timescales that were previously intractable. These advances, however, pose new challenges and will require new solutions to limitations inherent in such large-scale analyses. In this paper, we have shown that the particular combination of two independent advances; the construction of large, but incompletely resolved phylogenetic trees, and the assessment of phylogenetic conservatism in trait data, can bias inferences. We have provided one simple solution that might prove useful while efforts to resolve the Tree of Life are ongoing (see description *online*).<sup>7</sup>

#### ACKNOWLEDGMENTS

We thank C. Willis for sharing data on Thoreau's woods. This work was conducted as part of the Forecasting Phenology working group supported by the National Center for Ecological Analysis and Synthesis (NCEAS), a Center funded by NSF (DEB-0072909), the University of California–Santa Barbara, and the state of California. N. J. B. Kraft was supported by an NSERC CREATE training program in Biodiversity Research.

<sup>7</sup> <http://tolweb.org/tree/>

## LITERATURE CITED

- Ackerly, D. D. 2009. Phylogenetic methods in ecology. Encyclopedia of life sciences. John Wiley and Sons, Chichester, UK.
- Angiosperm Phylogeny Group. 1998. An ordinal classification for the families of flowering plants. *Annals of the Missouri Botanical Garden* 85:531–553.
- Angiosperm Phylogeny Group II. 2003. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG II. *Botanical Journal of the Linnean Society* 141:399–436.
- Angiosperm Phylogeny Group III. 2009. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Botanical Journal of the Linnean Society* 161:105–121.
- Blomberg, S. P., T. Garland, and A. R. Ives. 2003. Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution* 57:717–745.
- Cadotte, M. W., M. A. Hamilton, and B. R. Murray. 2009. Phylogenetic relatedness and plant invader success across two spatial scales. *Diversity and Distributions* 15:481–488.
- Cavender-Bares, J., K. H. Kozak, P. V. Fine, and S. W. Kembel. 2009. The merging of community ecology and phylogenetic biology. *Ecology Letters* 12:693–715.
- Crozier, R. H. 1997. Preserving the information content of species: genetic diversity, phylogeny, and conservation worth. *Annual Review of Ecology and Systematics* 28:243–268.
- Emerson, B. C., and R. G. Gillespie. 2008. Phylogenetic analysis of community assembly and structure over space and time. *Trends in Ecology and Evolution* 23:619–630.
- Faith, D. P. 1992. Conservation evaluation and phylogenetic diversity. *Biological Conservation* 61:1–10.
- Felsenstein, J. 1985. Phylogenies and the comparative method. *American Naturalist* 125:1–15.
- Grafen, A. 1989. The phylogenetic regression. *Philosophical Transactions of the Royal Society B* 326:119–157.
- Harmon, L. J., J. T. Weir, R. E. Glor, and W. Challenger. 2008. Geiger: investigating evolutionary radiations. *Bioinformatics* 24:129–131.
- Harvey, P. H., and M. D. Pagel. 1991. The comparative method in evolutionary biology. Oxford University Press, Oxford, UK.
- Hodkinson, T. R., and J. A. N. Parnell. 2007. Introduction to the systematics of species rich groups. Pages 3–20 in T. R. Hodkinson and J. A. N. Parnell, editors. *Reconstructing the tree of life: taxonomy and systematics of species rich taxa*. CRC Press, Boca Raton, Florida, USA.
- Kembel, S. W., P. D. Cowan, M. R. Helmus, W. K. Cornwell, H. Morlon, D. D. Ackerly, S. P. Blomberg, and C. O. Webb. 2010. Picante: R tools for integrating phylogenies and ecology. *Bioinformatics* 26:1463–1464.
- Kress, W. J., D. L. Erickson, A. F. Jones, N. G. Swenson, R. Perez, O. Sanjurjo, and E. Bermingham. 2009. Plant DNA barcodes and a community phylogeny of a tropical forest dynamics plot in Panama. *Proceedings of the National Academy of Sciences USA* 106:18621–18626.
- Losos, J. B. 2008. Phylogenetic niche conservatism, phylogenetic signal and the relationship between phylogenetic relatedness and ecological similarity among species. *Ecology Letters* 11:995–1007.
- Miller-Rushing, A. J., and R. B. Primack. 2008. Global warming and flowering times in Thoreau's Concord: a community perspective. *Ecology* 89:332–341.
- Paradis, E., J. Claude, and K. Srimmer. 2004. APE: an R package for analyses of phylogenetics and evolution. *Bioinformatics* 20:289–290.
- R Development Core Team. 2011. R. 2.13. R Project for Statistical Computing, Vienna, Austria.
- Revell, L. J., L. J. Harmon, and D. C. Collar. 2008. Phylogenetic signal, evolutionary process, and rate. *Systematic Biology* 57:591–601.
- Sibley, C. G., and J. E. Ahlquist. 1983. The phylogeny and classification of birds, based on the data of DNA-DNA hybridization. Pages 245–292 in R. F. Johnston, editor. *Current ornithology*. Plenum Press, New York, New York, USA.
- Sibley, C. G., and J. E. Ahlquist. 1987. Avian phylogeny reconstructed from comparisons of the genetic material, DNA. Pages 95–121 in C. Patterson, editor. *Molecules and morphology in evolution: conflict or compromise*. Cambridge University Press, London, UK.
- Sibley, C. G., and J. E. Ahlquist. 1990. *Phylogeny and classification of birds: a study in molecular evolution*. Yale University Press, New Haven, Connecticut, USA.
- Swenson, N. G., B. J. Enquist, J. Thompson, and J. K. Zimmerman. 2007. The influence of spatial and size scale on phylogenetic relatedness in tropical forest communities. *Ecology* 88:1770–1780.
- Uriarte, M., N. G. Swenson, R. L. Chazdon, L. S. Comita, W. J. Kress, D. Erickson, J. Forero-Montaña, J. K. Zimmerman, and J. Thompson. 2010. Trait similarity, shared ancestry and the structure of neighbourhood interactions in a subtropical wet forest: implications for community assembly. *Ecology Letters* 13:1503–1514.
- Vamosi, S. M., S. B. Heard, J. C. Vamosi, and C. O. Webb. 2009. Emerging patterns in the comparative analysis of phylogenetic community structure. *Molecular Ecology* 18:572–592.
- Webb, C. O., D. D. Ackerly, M. A. McPeck, and M. J. Donoghue. 2002. Phylogenies and community ecology. *Annual Review of Ecology and Systematics* 33:475–505.
- Webb, C. O., and M. J. Donoghue. 2005. Phylomatic: tree retrieval for applied phylogenetics. *Molecular Ecology Notes* 5:181–183.
- Wiens, J. J. 2008. Commentary: niche conservatism déjà vu. *Ecology Letters* 11:1004–1005.
- Wiens, J. J., et al. 2010. Niche conservatism as an emerging principle in ecology and conservation biology. *Ecology Letters* 13:1310–1324.
- Willis, C. G., B. Ruhfel, R. B. Primack, A. J. Miller-Rushing, and C. C. Davis. 2008. Phylogenetic patterns of species loss in Thoreau's woods are driven by climate change. *Proceedings of the National Academy of Sciences USA* 105:17029–17033.

## SUPPLEMENTAL MATERIAL

## Appendix

An alternate version of Fig. 2, assuming random-split starting trees with Grafen branch lengths, and a figure showing the relationship between  $K$  and tree size (*Ecological Archives* E093-023-A1).

## Supplement

R function to thin terminal polytomies (*Ecological Archives* E093-023-S1).