The Trouble with Topology: Phylogenies without Fossils Provide a Revisionist Perspective of Evolutionary History in Topological Analyses of Diversity

JAMES E. TARVER[1,2,*] AND PHILIP C. J. DONOGHUE[1,*]

[1]*School of Earth Sciences, University of Bristol, Wills Memorial Building, Queen's Road, Bristol, BS8 1RJ, UK and*
[2]*Department of Biological Sciences, Dartmouth College, Hanover, NH 03755, USA;*
Correspondence to be sent to: School of Earth Sciences, University of Bristol, Wills Memorial Building, Queen's Road, Bristol, BS8 1RJ, UK;
E-mail: james.tarver@bristol.ac.uk; phil.donoghue@bristol.ac.uk.

The decidedly asymmetrical architecture of the Tree of Life betrays the fact that some evolutionary lineages have diversified to a greater or lesser extent than have others. It has been a goal of evolutionary ecology to identify shifts in diversification rate (speciation rate minus extinction rate) and their causal bases. Historically, the question of relative diversity was addressed in terms of phylogenies of fossils (Simpson 1944) with focus shifting to taxonomies in the 1970s (Raup et al. 1973; Gould et al. 1977). The dramatic rise in the availability of molecular sequence data in the 1990s led to renewed interest in phylogenies and diversity from evolutionary biologists and attempts to explain patterns of diversity returned to phylogenies. For the first time, however, historical events inferred to have shaped current diversity were identified without recourse to palaeontological data (Harvey et al. 1994).

Two principal approaches have been developed– "distance" and "topological" methods. Distance (aka "temporal") methods represent a natural development of taxonomic methods that were initially aphylogenetic. They exploit branch length data or the temporal spacing of branching events within phylogenetic trees and while distance methods remain popular for analyzing rates of evolution, they have lost out in favor of topological methods for the analysis of the history of clade diversity.

Topological (aka "tree shape") methods were developed in parallel with distance methods but are distinguished in that they exploit only tree balance as a record of evolutionary history (Slowinski and Guyer 1989). They have been adopted widely because they eschew temporal and distance data and so they can be applied readily to phylogenetic trees, such as supertrees, where distance data are lacking (Moore et al. 2004). The purpose of these methods is to determine 1) if lineages within a given phylogenetic tree diversified under different rates and 2) which particular lineages within a given phylogenetic tree are more diverse than would be expected under a given model, such as a Yule model (Yule 1924). These methods were not, however, conceived to identify the causal bases of tree imbalance but merely to identify whether these phenomena might exist in the first place (Raup 1985; Slowinski and Guyer 1989). In this endeavour, topological methods

perform their task admirably. However, in practice, the overwhelming majority of studies (Table 1) that have employed topology-based methods have sought to identify causality "to know if shifts in diversification rate are correlated with changes in some other variable (e.g., the origin of morphological or behavioral novelties, ecological associations, biogeographic events)" (Moore et al. 2004; p. 524). Intrinsic causal factors are identified through coincidence with diversification rate shifts in tree topology and extrinsic causes are identified through their temporal correspondence to the age of the node on which a diversification rate shift is identified. Increasingly sophisticated methods are being developed to test the coincidence between the putative causal factors and the diversification rate shifts to which they are attributed (Moore and Donoghue 2007, 2009). However, the modus operandum is invariably inductive—first identifying diversification rate shifts and then seeking their causal bases rather than deductive—testing hypotheses of causal association between innovations and their impact upon diversity (but see Moore and Donoghue 2009).

Ultimately, hypotheses of causality rest on the intuitively reasonable assumption that the position of diversification rate shifts within the topology of a phylogenetic tree reflects their relative timing. Thus, diversification rate shifts identified near the root or the tips of a tree are considered to have occurred early or late, respectively, within the evolutionary history of the clade. In this contribution, we demonstrate that this assumption is not a natural expectation of phylogenetic trees of standing diversity. The topologies of census trees are in constant flux as taxa are added by speciation and pruned by extinction. Nonrandom imbalances in the diversity of sister clades need not be achieved by singular events or episodes of diversification, as topological methods presuppose. Nonrandom imbalances can also be achieved through temporally and causally unrelated episodes of random diversification (speciation and/or extinction), leading to spurious hypotheses of diversification rate shift. The contribution of extinction to standing diversity is especially critical, telescoping past flux in speciation and extinction into a single internal branch of a tree—an effect exploited in lineage through time analyses (Harvey et al. 1994). In this way, the summed

TABLE 1. All studies which use either the shift statistic (SymmeTREE) or the relative Cladogenesis statistic (End-Epi) to seek a causal explanation for perceived shifts in diversification rates. Additional papers have cited these two programs, however, they are either reviews or use other features implemented within said programs. Likewise, other methods (Rabosky et al. 2007; Alfaro et al. 2009) have been developed which are subject to the same limitations

Reference	Clade	Software used	Causal explanation sought
Steeman et al. 2009	Cetaceans	SymmeTREE	Palaeooceanographic changes
Kazancioglu et al. 2009	Parrotfishes	SymmeTREE	Sexual selection
Li et al. 2009	Schizothorax fishes	End-Epi	Geographic elevation
Egan and Crandall 2008	Flowering plants (Psoraleeae)	SymmeTREE	Climate change
Tolley et al. 2008	South African chameleons	SymmeTREE	Climate change
Merckx et al. 2008	Flowering plants (Burmanniaceae)	SymmeTREE	Climate change
Williams and Duda 2008	Gastropods	SymmeTREE and End-Epi	Tectonic activity
Hardman and Hardman 2008	Bullhead catfishes	SymmeTREE and End-Epi	Climate change
Ruta et al. 2007	Temnospondyls	SymmeTREE	Recovery after mass extinctions
Moore and Donoghue 2007	Flowering plants (Adoxaceae and Valerianaceae)	SymmeTREE and End-Epi	Biogeographic and morphological changes
Verdu et al. 2007	Multiple plant clades	SymmeTREE	Life history traits
Forest et al. 2007	Milkworts	SymmeTREE	Morphological novelty
Wiens et al. 2007	Tropical salamanders	End-Epi	Geographic elevation
McLeish et al. 2007	Acacia thrips	SymmeTree	Life history traits
Wiens et al. 2006	North American salamanders	End-Epi	Geographic elevation
McKenna and Farrell 2006	Leaf beetles	SymmeTREE and End-Epi	Climate change
Davison and Chiba 2006	Land snails	End-Epi	Adaptive radiation
Kozak et al. 2006	North American salamanders	End-Epi	Climatic change
Jones et al. 2005	Bats	SymmeTREE	Coevolution with flowering plants
Kozak et al. 2005	Dusky salamanders	End-Epi	Life history traits
Near and Benard 2004	Logperch darters	End-Epi	Sea level fluctuations
Jennings et al. 2003	Australian lizards	End-Epi	Climate change
Ruber et al. 2003	American gobies	End-Epi	Behavior
Vences et al. 2002	Malagasy treefrogs	End-Epi	Life history traits
Bininda-Emonds et al. 1999	Carnivores	End-Epi	Adaptive radiation
Fumagalli et al. 1999	Shrews	End-Epi	Chromosomal variation
Cook and Lessa 1998	Rodents	End-Epi	Life history traits
Price et al. 1998	Warblers	End-Epi	Climate change

effects of random processes can produce nonrandom tree topologies.

If the veracity and timings of diversification rate shifts are to be correctly identified, we argue that it is necessary to distinguish the relative timing of the contributory episodes of speciation and extinction. In achieving this objective, the architecture of trees must be considered in terms of extinct and not merely extant taxa. In effect, it is necessary to shift focus from the investigation of diversification rate, to speciation rate, which better fits the pure birth Yule model that underpins the majority of topology-based methods. Only once this has been established will it be possible to speculate on, and test among, putative causal bases underpinning the flux of tree shape over the evolutionary history of clades.

Materials and Methodology

Using empirical data sets, we demonstrate the effect of extinction upon tree topology, perceived shifts in diversification rate, and the relative timing of the action of extinction versus its perceived impact on a tree. We then show how the integration of extinct among extant members of a clade makes it possible to distinguish between the contributory effects of speciation and extinction to diversity, to distinguish genuine from artifactual diversification rate shifts, and establish the timing of their causal drivers. For our demonstration, we used two empirical data sets, a supertree of extant carnivore relationships (Bininda-Emonds et al. 2007) to examine the impact of extinction, and a phylogenetic tree of extant and extinct crocodilian relationships to examine the impact of speciation (Gatesy et al. 2004). Our focus is on the identification of diversification rate shifts because, in terms of the relative timing of the events they predict, they are more readily testable than other topology-based methods that, for instance, focus exclusively on tree symmetry. Of the two approaches most commonly used to identify diversification rate shifts, End-Epi (Rambaut et al. 1997) is no longer distributed and runs on an obsolete Macintosh operating system and SymmeTREE (Chan and Moore 2005). SymmeTREE has been adopted broadly, and we use it as an exemplar of the performance of topology-based methods.

Although topological methods seek to identify shifts in diversification rate, they do not measure rates in units of absolute time. Rather, topological methods assess whether local subtrees are more unbalanced than would be expected by chance if rates of speciation were equal. When analyses are run in SymmeTREE, two shift statistics Δ^1 and Δ^2 are produced for each node within the phylogenetic tree. A P value is also calculated for each node to express the probability of the observed difference in sister-group diversity. Significant ($P < 0.05$) and substantial ($P < 0.1$) values indicating asymmetries in sister clade diversity that reject the equal rates Markov (ERM) null model indicating that the more speciose clade has undergone nonrandom diversification. If two clades are considered, the shift statistics look at the

probability that a shift occurred on the internal branch that subtends the more speciose clade. The Δ^1 shift statistic is simply the difference in the log-likelihood ratios between two models, one with a homogeneous rate, and the other with a heterogeneous rate at both the basal and the internal nodes of the clade (a,(b,c)). The Δ^2 shift statistic selectively and systematically excludes ingroup subclades in estimating the net rate of diversification in order to pinpoint nodes where the rate may have changed radically (see Chan and Moore 2002, 2005; Moore et al. 2004). The specific settings used in our analyses were 1,000,000 random resolutions of the entire tree and 10,000 for nonbifurcating nodes using the taxon-size sensitive ERM algorithm, and the Δ^2 shift statistic was used to identify diversification rate shifts.

Topologies of Diversification Rate Shifts do not Reflect Timing

SymmeTREE analysis of the Carnivora phylogenetic tree identified three significant and one substantial shift in the rate of diversification, all at levels approximately midway along the temporal axis (Fig. 1). Molecular clock dates for these lineage divergences suggest that these shifts in the rate of diversification occurred in the interval 13–50 Ma (Bininda-Emonds et al. 2007). To demonstrate how the topological position of diversification phenomena flux with extinction, we simulated the impact that extinction would have on the perspective of evolutionary history for extant carnivorans. Taxa that are considered "critically endangered" or "endangered" in the latest available International Union for Conservation of Nature (IUCN) and the United States Endangered Species Act (US ESA) lists (Released 1st September 2007) were pruned from the phylogenetic tree, representing a pessimistic census date perhaps 100 years hence. This is an entirely arbitrary but objective filter to simply demonstrate the effect of extinction upon the perception of historical diversity and, thus, perceived shifts in the rate of diversification.

The removal of 31 species (12%) leads to substantially different perception of diversification over what is essentially the same episode of evolutionary history. Although some shifts are identified in both topologies, others occur in only one of the topologies (Fig. 1). For example, a shift leading to the families Procyonidae and Mustelidae is lost in the pruned tree (Fig. 1L), and a new shift appears in the lineage leading to the clade encompassing Procyonidae, Mustelidae, and Pinnipedia (Fig. $1E_1$). This greatly affects the timing of the putative events that may have led to either diversification rate shift, with the dating of this phenomenon pushed back from 38 to 50 Ma. Likewise, a new diversification rate shift is identified as a result of taxon pruning at the base of the Civet and Genet clade (Fig. $1E_2$). The split between this clade and its sister taxon occurred at approximately 24 Ma during a period of global warming following a global decrease in temperatures after the Eocene–Oligocene extinction event (Cowie 2007).
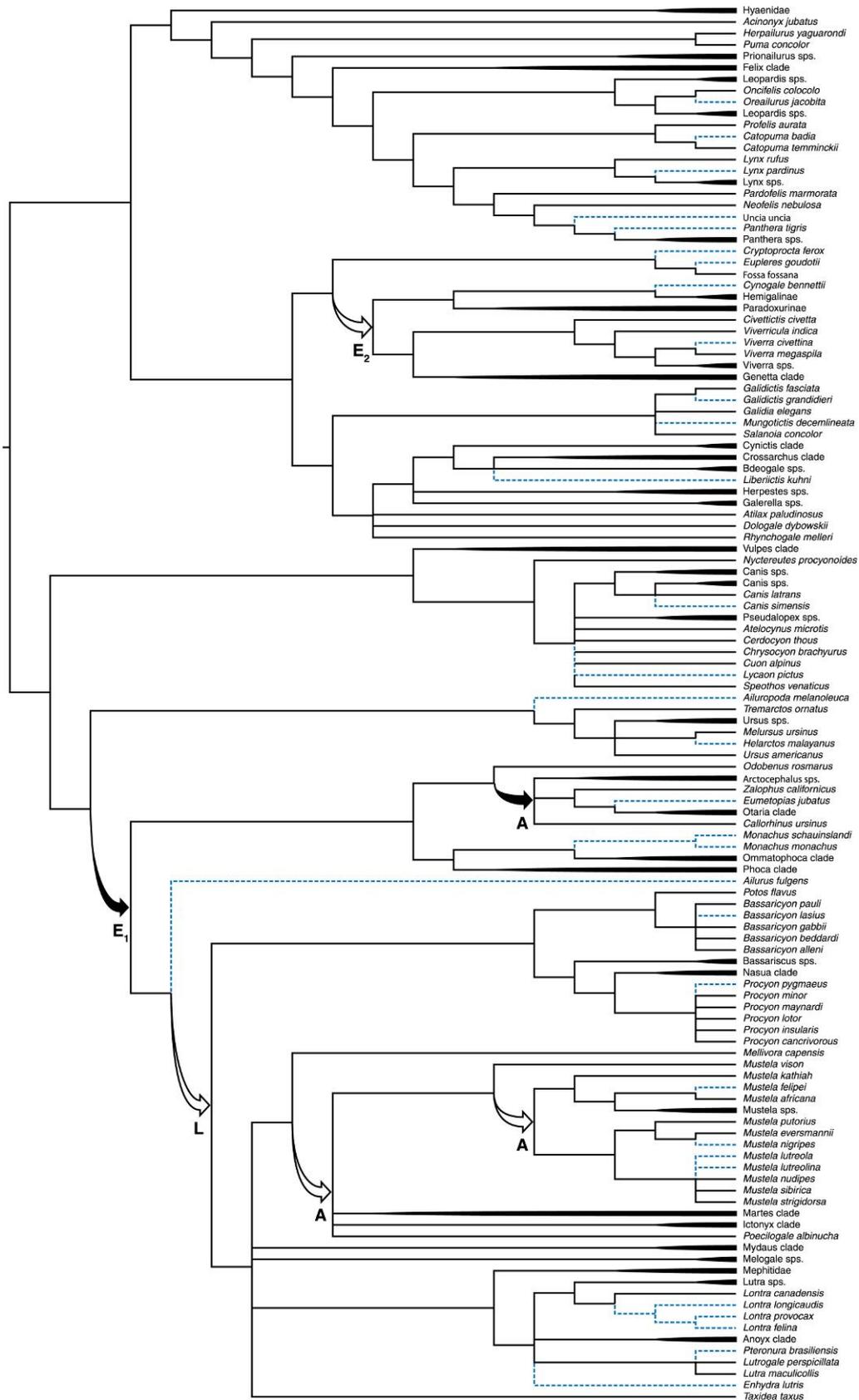
Critically, however, the causal factor of this putative diversification rate shift is not rooted in a narrative of global warming in the Oligocene, but prospective extinction resulting from poaching and habitat destruction by humans in the Holocene.

Evidently, the positions of putative diversification rate shifts in phylogenetic trees do not reflect their relative or absolute timing. Diversification rate shifts located deep within a phylogenetic tree can be driven by random or deterministic causal factors operating at any time between the present and the node on which the rate shift is identified. Further, diversification rate shifts need not represent "events," as they are usually portrayed but instead may result from the net effects of multiple rounds of stochastic and/or deterministic speciation and/or extinction. This occurs because, as taxa are removed from consideration when they become extinct, the resolution of perceived changes in diversification rate is diminished. In some instances, this will mask real shifts in diversification rate, whereas in other instances, it will lead to the identification of spurious diversification rate shifts. As a result, the identification of diversification rate shifts may be little more than sampling artefact. The allusion to any deterministically causal macroevolutionary process is entirely inferential. No topology-based analyses of diversification rate have entertained this variable and so, for this reason, their results must be considered questionable.

Constraining the Timing of Diversification Rate Shifts Requires the Inclusion of Extinct Taxa

Irrespective of the performance of topology-based methods in identifying diversification rate shifts, the phenomenon that we wish to explain, the unequal distribution of species diversity is real, and it remains our objective to find an approach that allows us to identify the cause of imbalances in clade structure. Because the identification of causal factors underpinning diversification rate shifts relies upon little more than the weak inductive logic of temporal correlation, it is imperative that there is constraint on the timing of both the diversification rate shifts and their putative causes (Moore and Donoghue 2009). The timing of action of the causal factors (deterministic or stochastic) that gave rise to shifts in taxic richness can only be determined by distinguishing the contributory effects of speciation and extinction. This requires that topological analyses of extant diversity also consider the effect of cumulative random extinction as a causal explanation in addition to deterministic factors. However, rejecting

such a hypothesis requires the placement of extinct taxa within such topologies to be considered. Ultimately, the goal of topology-based methods cannot be achieved by analyzing phylogenetic trees that encompass extant taxa alone—"complete" phylogenies are required that encompass not just extant members of the clade in question but also their extinct relatives.

Concern over taxon sampling in topological analyses was raised by Mooers (1995) who showed that incomplete trees are more imbalanced than complete trees. This is significant because an increase in imbalance may lead to the erroneous identification of diversification rate shifts (Heath et al. 2008). It has been argued that the imbalance between complete and incomplete phylogenies reflects the nonrandom manner by which systematists sample taxa; if a random sample were taken, the results for the incomplete and complete phylogenies would be the same (Guyer and Slowinski 1991; Kirkpatrick and Slatkin 1993; Mooers 1995; Purvis and Agapow 2002). However, more recent work has shown that the random removal of taxa from simulated complete phylogenies increases the observed imbalance and, thus, incomplete sampling can have a strong effect on the perception of macroevolutionary events inferred from tree topologies (Heath et al. 2008). This provides further support for the inclusion of extinct taxa among their extant relatives in topological analyses of macroevolutionary processes such as diversification.

Extinct Taxa and Topology-Based Methods

From their conception, it has been argued that topology-based methods should not be applied to extinct taxa:

> When extinct lineages are used, relative differences in extinction can no longer be tested; instead, only differential speciation can be considered. However, if significant differences are found, they could be the results of certain lineages' having slightly more extinct taxa than others and, therefore, fewer chances through time for speciation. (Slowinski and Guyer 1989; p. 910)

It has also been suggested that extinct taxa should not be included within topology-based methods of analyzing diversity because all such methods, including SymmeTREE (Chan and Moore 2005), End-Epi (Rambaut et al. 1997), and apTreeshape (Bortolussi 2006) contrast the degree to which the relative diversity of sister clades deviates from a pattern of random

FIGURE 1.   Carnivoran phylogenetic tree taken from Bininda-Emonds et al. (2007). The SymmeTREE analysis was run on both this tree and a pruned version, with all taxa at risk from extinction removed (dashed lines). The results of both analyses are shown with letters above each node corresponding to the presence or absence of a diversification rate shift in the respective analyses. Nodes represented by filled arrows indicate significant shifts in diversification, unfilled arrows indicate substantial shifts. (A) corresponds to shifts present in both phylogenies, (E) corresponds to shifts created by the future extinction of taxa, whereas (L) corresponds to the loss of a shift by extinction. Both represent artifacts of the data. Black triangles indicate the collapse of a larger clade for ease of representation.

diversification, provided by a Yule Model (Kirkpatrick and Slatkin 1993; Fusco and Cronk 1995; Mooers and Heard 1997; McKenzie and Steel 2000; Chan and Moore 2002; Heard and Mooers 2002; Blum and Francois 2005). A Yule model is a pure birth model (Yule 1924; Aldous 2001; Nee 2006) that follows a random branching process in which each tip (taxon) has an equal probability of branching at any moment in time. Clearly, speciation is not equally probable in extinct and extant lineages and, hence, extinct taxa have either been actively pruned from phylogenetic trees before conducting analyses (e.g., Jones et al. 2005) or else omitted in the first instance (e.g., Bininda-Emonds et al. 2007).

Existing software used in analyzing diversification rate requires users to exclude extinct taxa because of the pure birth algorithms that were coded into them, but this is not a general requirement for software implementations of topology-based methods. Previously, analyses of tree topology have used statistics that have incorporated variable rates of speciation (Heard 1996), speciation followed by mass extinction events (Yule 1924; Heard and Mooers 2002), and variable rates of both speciation and extinction (Raup 1985; Harcourt-Brown et al. 2001; Harcourt-Brown 2002). To realize the aim of identifying the causal bases of diversification rate shifts it is necessary to distinguish the contributory effects of speciation versus extinction. This can only be achieved readily by analyzing holistic phylogenies that include extinct taxa.

Time Slicing does not Resolve the Problems with Inclusion of Extinct Taxa

A solution to the problem of including extinct taxa in topology-based analyses of diversity was proposed and applied by Ruta et al. (Ruta et al. 2007; Lloyd et al. 2008), but their approach fails to overcome the fundamental problem with topology-based approaches, viz. an inability to distinguish between the relative effects of speciation and extinction. Their method also creates several new problems concerning the interpretation of results.

Ruta et al. (2007) argued that topology-based methods may be extended to include extinct taxa without violating the Yule model by considering tree topology in cohorts of contemporaries. Once the phylogenetic tree has been mapped onto stratigraphy, it is broken down into separate time slices on which topology-based analyses are conducted sequentially from the oldest to the youngest. When an extinct taxon falls outside the temporal extent of this analytical window, it is removed. Lineages that are inferred to have existed within the time interval because subsequent intervals include descendent clades are shown as ghost lineages (Fig. 2). This approach overcomes the violation of the particular implementation of the Yule model in SymmeTREE (Chan and Moore 2005) because within each time slice every taxon is extant. It also overcomes the limitation of topology-based methods in that it provides a basis for including extinct taxa alongside extant lineages. However, because this method considers tree topology within successive temporal intervals, it remains a topological approach and subject to the same errors associated with the analysis of extant taxa alone, with workers unable to distinguish between the relative effects of both speciation and extinction.

To demonstrate this phenomenon, we utilized the crocodilian phylogenetic tree of Gatesy et al. (2004) mapped onto geological time (Fig. 3) using data from Brochu (1997, 1999, 2000, 2003). In so doing, we acknowledge that this phylogenetic tree is contentious and that the temporal ranges of these lineages will undoubtedly be revised in light of further sampling of the fossil record and molecular clock analyses. However, none of the competing phylogenies encompass so many extant and extinct taxa and the stratigraphic range data are constant in our comparative analyses of diversification rates. Therefore, with these caveats, we use these resources to make our conceptual point; the results of
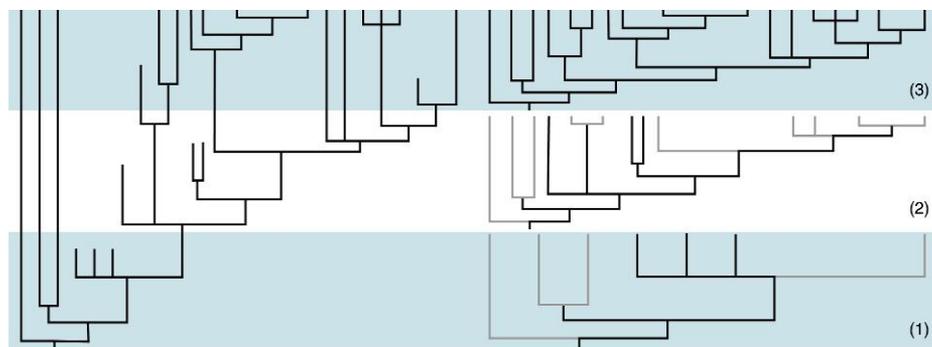

FIGURE 2. The time slice approach as advocated by Ruta et al. (2007). The complete phylogenetic tree is shown on the left, whereas the 3 respective time slices are shown on the right. Each slice shows a different portion of the evolutionary history of the clade. Slice one shows the origination of the clade with the gray line representing ghost ranges of clades that have yet to arise and are not currently known in the fossil record. The black lines indicate taxa with a known fossil record. In slice two, the three taxa known from slice one have become extinct and so no longer included within the analysis, whereas some of the ghost ranges have led to taxa now known in the fossil record others are still present. In the third and final slice, all taxa are represented as present and no ghost ranges are shown but several taxa from slice two have been removed because of extinction.
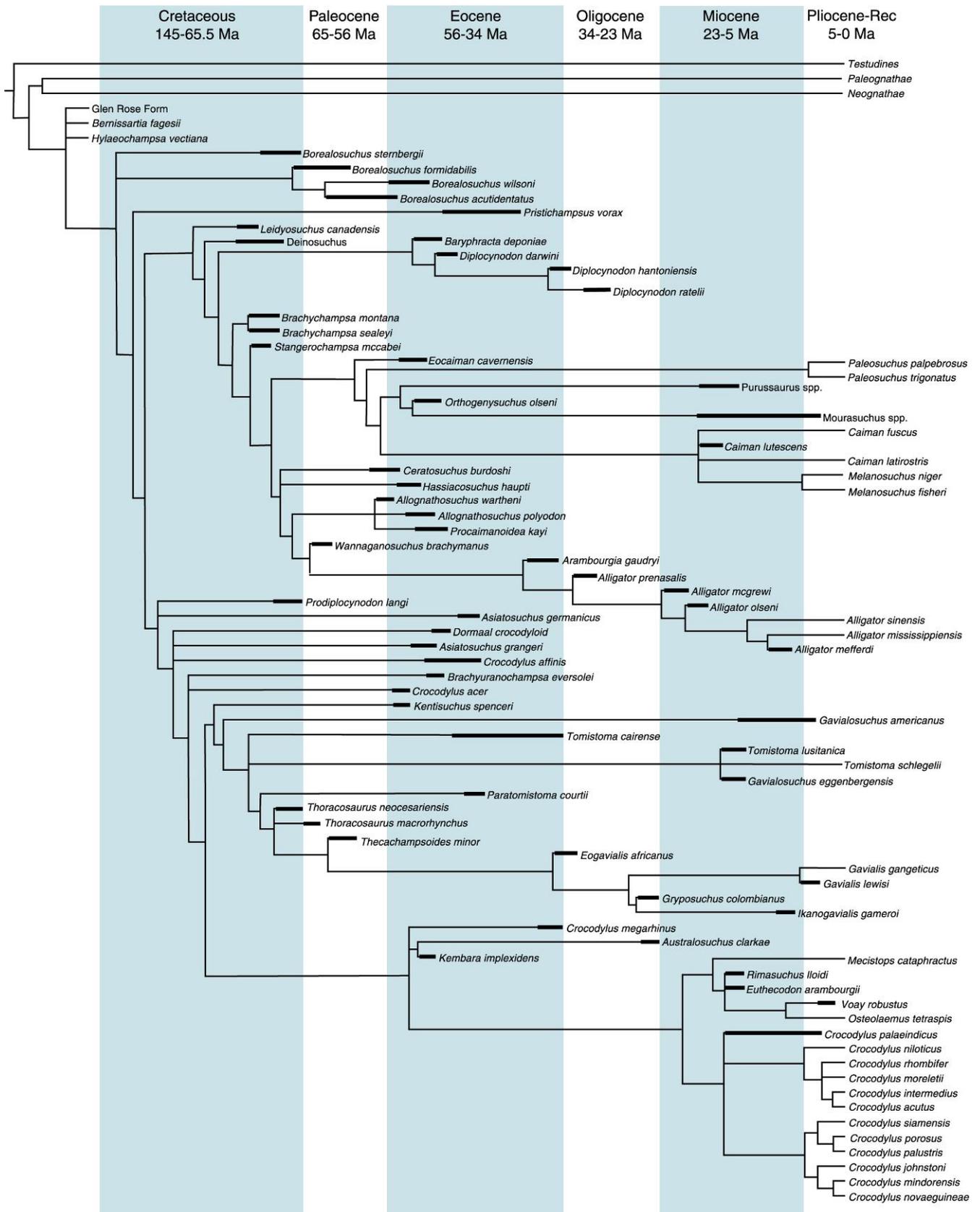
FIGURE 3. The crocodilian phylogenetic tree as presented in Gatesy et al. (2004) mapped to stratigraphy and dated according to Brochu (1997, 1999, 2000, 2003).

our analyses should not be read as providing insight into crocodilian evolutionary history.

The method advocated by Ruta et al. (2007) was applied to 6 time slices (Fig. 4). No shifts were observed within the Pliocene, Miocene, and Oligocene. There was one shift in both the Eocene and the Cretaceous, whereas 6 were observed during the Paleocene. How to interpret the results? Should only those shifts present in all time slices be considered? Those shifts that are present in the majority? Or are all shifts identified in all analyses equally valid regardless of their transience? Another option would be to consider the time slices themselves: 6 shifts are observed in the Paleocene slice of which 5 are unique to this slice. If this were the correct approach, then the high number of diversification rate shifts in the Paleocene (Fig. 4) would imply a deterministic factor, possibly associated with the recovery following the Cretaceous–Palaeogene mass extinction. However, when these shifts are considered in light of their stratigraphic occurrence, it is clear to see that they all occurred at least 10 myr before the Cretaceous–Palaeogene mass extinction.

The underlying problem with the method advocated by Ruta et al. (2007) is that, like topological analyses of extant taxa, it relies upon incomplete phylogenies. As per topological analyses of exclusively extant taxa, the solution is to discriminate the relative effects of speciation and extinction on diversity by analyzing speciation alone and entertaining extinction as causal factor in explanation of lineages that are more diverse than would be expected under the Yule Model.

A Solution: Analysis of Speciation Rate

Although topology-based methods perform their task well, they cannot discriminate between stochastic and deterministic differences in diversity or the relative timing of shifts in diversity. However, it remains possible to achieve the principal aims underpinning topology-based analyses of diversity. This can be done by discriminating between changes in diversity that are genuinely the outcome of contemporaneous extinction and speciation from changes in diversity that are merely the outcome of telescoping multiple episodes of stochastic extinction.

These aims can be achieved by analyzing speciation alone and attempting to distinguish between imbalances in diversity that deviate from the spectrum of outcomes expected on the basis of the Yule model. This is the approach advocated for analysis of trees encompassing extinct taxa when topology-based methods of diversity were first proposed (Slowinski and Guyer 1989: p. 910). It is a logical expectation that, unlike analyses of diversification, shifts in the speciation rate are contemporaneous with the relative age of the branch on which the rate shift is identified. It is necessary to consider extinction as an explanatory factor, among other possible intrinsic and extrinsic factors. As we have shown, however, this was always an implicit

requirement of effective interpretations of diversification rate shifts. This approach also has the benefit of allowing—arguably requiring—the inclusion of extinct taxa, facilitating discrimination between apparent and actual historical diversities (cf. Harvey et al. 1994).

Rather than applying a time slicing approach, we used a nested-growth method wherein topology-based analytical methods are applied to increasingly inclusive nested sets of taxa, from the oldest to the youngest taxon. In this instance, the phylogenetic tree is grown through time so that all speciation events are recorded, and the SymmeTREE analysis is repeated after the addition of each successive time interval (i.e., the first analysis is on time period T1, the second is on T1 + T2 and the third on T1 + T2+ T3). This makes it possible to discriminate the pattern of speciation alone to which the temporally varying pattern of diversification may be compared. In these analyses, taxa move into, but not out of, the analytical reading frame of an analysis of tree topology (Fig. 5). Thus, tree topology changes only as a result of speciation not extinction. Anomalous shifts in diversification rates that arise as a result of the inclusion of extinct taxa are identifiable because they appear at the base of extant clades long after their fossil relatives have gone extinct (cf. Slowinski and Guyer 1989).

In the crocodilian example, Shift 4 can be identified as the product of extinction rather than speciation (Fig. 5). Although the shift is inferred to have occurred in the Late Cretaceous, it is not apparent until the reading frame is extended to the Miocene. In this case, the diversification rate shift was caused by two factors; the later speciation within the Crocodylinae during the Miocene and the extinction of numerous (now stem) taxa during the Eocene. These two events occurred approximately 50 and 20 myr after the shift is inferred to have occurred showing that the causative factors were not associated with the origination of the more speciose clade.

This is in contrast to Shift 1 that is present in every reading frame from the Cretaceous to the present. This shift is identified while both clades are contemporaneous, with the less speciose clade surviving more than 35 myr after the shift is inferred to have occurred showing that this lineage had sufficient longevity for further speciation. These results indicate that the shift at the base of crown Crocodylia reflects a genuine event during the end Cretaceous rather than an artifact caused by later extinction or speciation.

Caveat Ex Tempore

Our approach to using topology-based methods to identify clades that are more speciose than would be anticipated under a pure birth Yule model has the benefit of effectively identifying the relative timing of shifts in speciation. If only extant taxa are considered, it is not possible to discriminate whether extinction (stochastic or deterministic) is a causal factor in explaining differences in diversity, nor indeed, the relative timing of extinction or any other causal factor. However, with the inclusion of extinct taxa and subsequent conversion
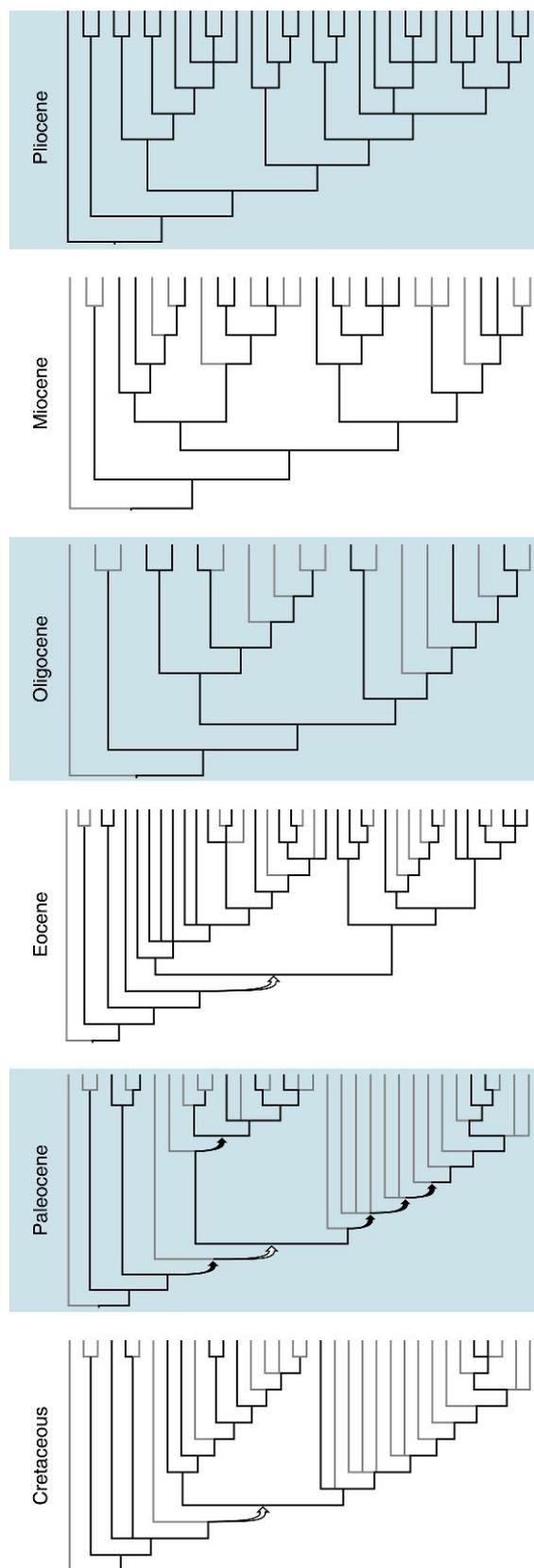
FIGURE 4.   The crocodylian phylogenetic tree divided into the 6 time slices shown in Fig. 3 following the method of Ruta et al. (2007). Black lines indicate taxa present within the time slice, whereas gray lines show taxa known from their ghost ranges. Nodes represented by filled arrows indicate significant shifts in diversification, unfilled arrows indicate substantial shifts.
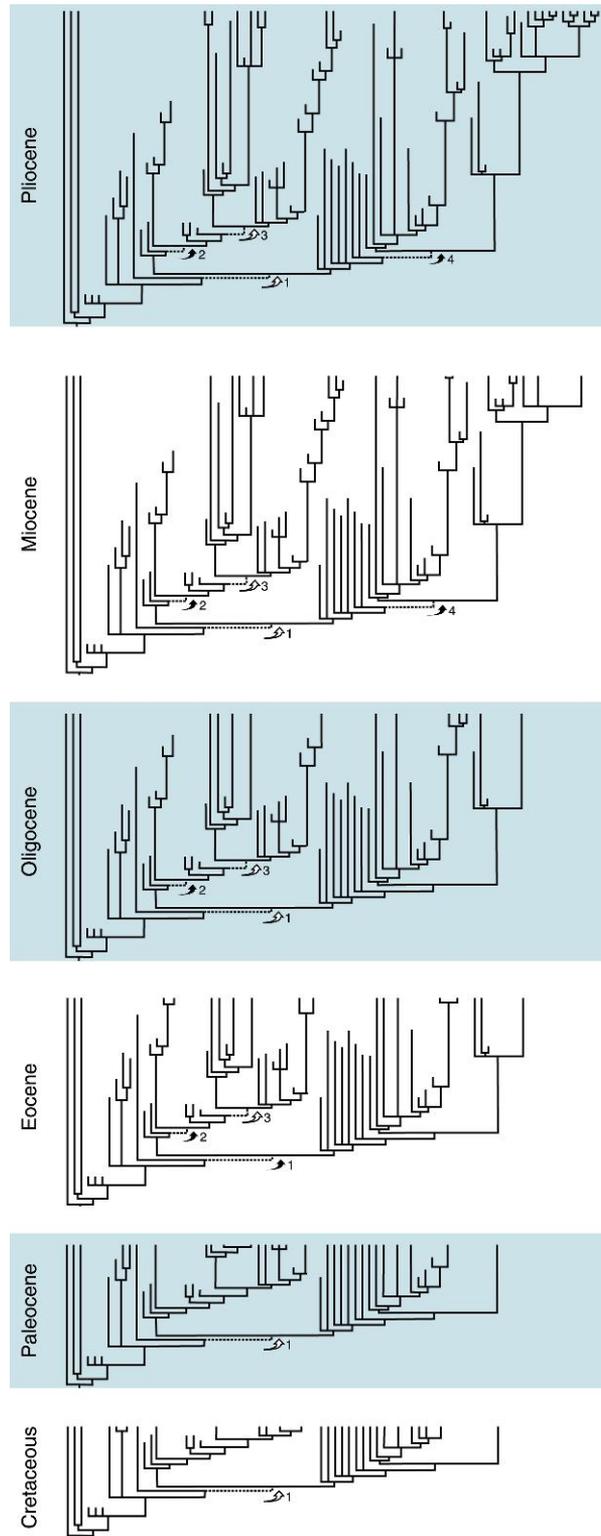
FIGURE 5. The Gatesy et al. (2004) crocodilian phylogenetic tree analyzed using the nested-growth method. The phylogenetic tree is grown through the 6 time periods, nodes represented by filled arrows indicate significant shifts in diversification, and unfilled arrows indicate substantial shifts. As the tree is grown through time, the relative effects of both speciation and extinction becomes clear and it is possible to distinguish between the two. When we consider the diversification rate shift labeled 1, we see that it is present in every time period. It appeared in the Cretaceous, while both lineages were extant, the extinction of Pristichampsus vorax during the Eocene would not have had any affect on this diversification rate shift as it was already present when both lineages were contemporaneous. This is in contrast to the shift labeled 4, this shift appeared in the Miocene, although it is inferred to have occurred in the Cretaceous and is an artifact caused by both the speciation of the Crocodylinae in the Miocene and the extinction of many stem taxa during the Eocene.

to timetrees (whether by using the temporal data with which fossil taxa are intrinsically imbued or else relying on molecular clock estimates), it is possible to discriminate between stochastic and nonrandom extinction as a causal factor underpinning diversity, as well as the relative, even absolute, timing of nonstochastic shifts in diversity.

Despite our enthusiasm for the integration of extinct and extant taxa in topology-based analyses of diversity, the view that such a holistic approach in some way provides for "complete" phylogenies must be tempered with the realization that it will never be possible to obtain phylogenies that encompass all members of a clade. This is because only a small proportion of species preserve a fossil record that survives the rigors of tectonism and is recovered and identified as relevant (Paul 2009). Further, the temporal extent of those extinct species that are known to science will always fall short of their true longevity because of secular biases in the preservation of environments in the rock record, peculiarities in the way that fossils are sampled, and the impact of provincialism in their biogeographic distribution, among other factors (e.g., Donoghue and Benton 2007). These limitations can be constrained somewhat using methods such as confidence intervals (Marshall 1997), and graphic correlation (Mann and Lane 1995) on stratigraphic range data, as well as molecular clock analyses (Smith et al. 2006). Indeed, these limitations may be overcome altogether if preservation is controlled for alongside estimates of speciation and extinction (Connolly and Miller 2001; Foote 2001; Alroy et al. 2008), but no methods yet exist that integrate all three variables in a tree-based approach. Regardless, the temporal context of those species that are preserved, collected, and studied, should provide sufficient constraint to discriminate between stochastic extinction and deterministic causal factors in explanation of nonstochastic patterns of clade diversity. The temporal correlation of putative causal factors may then be tested probabilistically (Moore and Donoghue 2009).

At the conception of topology-based methods, Slowinski and Guyer (1989) shyed away from analyses that encompass extinct taxa, perhaps implicitly because of the problems we have just outlined but explicitly because of the problems perceived with identifying the phylogenetic position of fossil taxa. Empirical evidence shows that extinct taxa are just as influential as their extant relatives in affecting the precision of phylogenetic analyses (Gauthier et al. 1988; Donoghue et al. 1989; Cobbett et al. 2007) and simulations show in attempting to achieve an integrated phylogenetic tree of extinct and extant taxa, molecular phylogenetic data greatly improves the accuracy with which the phylogenetic position of extinct taxa is identified (Wiens 2009).

Finally, we return to the nub of the issue, that is, why are certain clades considerably more speciose than their immediate relatives? Answering this question has been one of the principal goals of biology and, if it is to be achieved, it will be necessary to overcome both the limitations of the fossil record and the negative prejudices that exist toward it. This echoes the call for the integration of fossil taxa into lineage through time (Harvey et al. 1994; Pybus and Harvey 2000) based analyses of historical diversity (Crisp and Cook 2009; Rabosky 2009, 2010a,b; Liow et al. 2010; Quental and Marshall 2010). Ultimately, the fossil record may not be complete but, then, neither is a molecular phylogeny that, by necessity, encompasses only extant taxa. The only means of approximating a complete phylogeny and, therefore, a complete perspective on historical diversity is by restoring extinct taxa to their rightful place in phylogenies alongside their living relatives (Wiens et al. 2010).

CONCLUSIONS

Topology-based methods provide a means of describing tree shape and determining whether or not the shape of a particular phylogenetic tree deviates from the kind of shape that would be produced by a homogenous process of random branching. Inevitably, where differences in the diversity of sister lineages deviate significantly from random chance, causal drivers are sought. However, phylogenetic tree topology is the sum of multiple rounds of speciation and extinction, resulting in tree shapes that are significantly nonrandom. Furthermore, the relative timing of causal drivers underpinning tree asymmetries need not relate to the position of the root of the asymmetry within the topology of the tree. To determine the material basis of tree asymmetries, it is necessary to discriminate between those that occur as a result of contemporaneous changes in diversification rates versus tree asymmetries that occur through extinction-driven loss of phylogenetic history. This cannot be achieved using topological data from extant species alone. Fossil taxa incorporated into phylogenies of their extant relatives provide a means of temporally constraining and discriminating between the relative contributions of speciation and extinction in effecting diversification.

REFERENCES

Aldous D.J. 2001. Stochastic models and descriptive statistics for phylogenetic trees, from Yule to today. Stat. Sci. 16:23–34.
Alfaro M.E., Santini F., Brock C., Alamillo H., Dornburg A., Rabosky D.L., Carnevale G., Harmon L.J. 2009. Nine exceptional radiations

plus high turnover explain species diversity in jawed vertebrates. Proc. Natl. Acad. Sci. U.S.A. 106:13410–13414.

Alroy J., Aberhan M., Bottjer D.J., Foote M., Fursich F.T., Harries P.J., Hendy A.J.W., Holland S.M., Ivany L.C., Kiessling W., Kosnik M.A., Marshall C.R., McGowan A.J., Miller A.I., Olszewski T.D., Patzkowsky M.E., Peters S.E., Villier L., Wagner P.J., Bonuso N., Borkow P.S., Brenneis B., Clapham M.E., Fall L.M., Ferguson C.A., Hanson V.L., Krug A.Z., Layou K.M., Leckey E.H., Nurnberg S., Powers C.M., Sessa J.A., Simpson C., Tomasovych A., Visaggi C.C. 2008. Phanerozoic trends in the global diversity of marine invertebrates. Science. 321:97–100.

Bininda-Emonds O.R.P., Cardillo M., Jones K.E., MacPhee R.D.E., Beck R.M.D., Grenyer R., Price S.A., Vos R.A., Gittleman J.L., Purvis A. 2007. The delayed rise of present-day mammals. Nature. 446: 507–512.

Bininda-Emonds O.R.P., Gittleman J.L., Purvis A. 1999. Building large trees by combining phylogenetic information: a complete phylogeny of the extant Carnivora (Mammalia). Biol. Rev. 74:143–175.

Blum M.G.B., Francois O. 2005. On statistical tests of phylogenetic tree imbalance: The Sackin and other indices revisited. Math. Biosci. 195:141–153.

Bortolussi N., Durand E., Blum M., Francois O. 2006. apTreeshape: statistical analysis of phylogenetic tree shape. Bioinformatics. 22: 363–364.

Brochu C.A. 1997. Morphology, fossils, divergence timing, and the phylogenetic relationships of *Gavialis*. Syst. Biol. 46:479–522.

Brochu C.A. 1999. Taxon sampling and reverse successive weighting. Syst. Biol. 48:808–813.

Brochu C.A. 2000. Phylogenetic relationships and divergence timing of Crocodylus based on morphology and the fossil record. Copeia. 3:657–673.

Brochu C.A. 2003. Phylogenetic approaches toward crocodylian history. Annu. Rev. Earth Planetary Sci. 31:357–397.

Chan K.M.A., Moore B.R. 2002. Whole-tree methods for detecting differential diversification rates. Syst. Biol. 51:855–865.

Chan K.M.A., Moore B.R. 2005. SYMMETREE: whole-tree analysis of differential diversification rates. Bioinformatics. 21:1709–1710.

Connolly S.R., Miller A.I. 2001. Joint estimation of sampling and turnover rates from fossil databases: capture-mark-recapture methods revisited. Paleobiology. 27:751–767.

Cobbett A., Wilkinson M., Wills M.A. 2007. Fossils impact as hard as living taxa in parsimony analyses of morphology. Syst. Biol. 56: 753–766.

Cook J.A., Lessa E.P. 1998. Are rates of diversification in subterranean South American tuco-tucos (genus Ctenomys, Rodentia: Octodontidae) unusually high? Evolution. 52:1521–1527.

Cowie J. 2007. Climate change: biological and human aspects. 1st ed. Cambridge: Cambridge University Press.

Crisp M.D., Cook L.G. 2009. Explosive radiation or cryptic mass extinction? Interpreting signatures in molecular phylogenies. Evolution. 63:2257–2267.

Davison A., Chiba S. 2006. Labile ecotypes accompany rapid cladogenesis in an adaptive radiation of Mandarina (Bradybaenidae) land snails. Biol. J. Linn. Soc. 88:269–282.

Donoghue P.C.J., Benton M.J. 2007. Rocks and clocks: calibrating the tree of life using fossils and molecules. Trends Ecol. Evol. 22: 424–431.

Donoghue, M.J., Doyle J., Gauthier J., Kluge A., Rowe T. 1989. The importance of fossils in phylogeny reconstruction. Annu. Rev. Ecol. Syst. 20:431–460.

Egan A.N., Crandall K.A. 2008. Divergence and diversification in North American Psoraleeae (Fabaceae) due to climate change. BMC Biol. 6. doi:10.1186/1741-7007-6-55.

Foote M. 2001. Inferring temporal patterns of preservation, origination, and extinction from taxonomic survivorship analysis. Paleobiology. 27:602–630.

Forest F., Chase M.W., Persson C., Crane P.R., Hawkins J.A. 2007. The role of biotic and abiotic factors in evolution of ant dispersal in the milkwort family (Polygalaceae). Evolution. 61:1675–1694.

Fumagalli L., Taberlet P., Stewart D.T., Gielly L., Hausser J., Vogel P. 1999. Molecular phylogeny and evolution of Sorex shrews (Soricidae: Insectivora) inferred from mitochondrial DNA sequence data. Mol. Phylogenet. Evol. 11:222–235.

Fusco G., Cronk Q.C.B. 1995. A new method for evaluating the shape of large phylogenies. J. Theoret. Biol. 175:235–243.

Gatesy J., Baker R.H., Hayashi C. 2004. Inconsistencies in arguments for the supertree approach: supermatrices versus supertrees of Crocodylia. Syst. Biol. 53:342–355.

Gauthier J., Kluge A.G., Rowe T. 1988. Amniote phylogeny and the importance of fossils. Cladistics. 4:105–209.

Gould S.J., Raup D.M., Sepkoski J.J., Schopf T.J.M., Simberloff D.S. 1977. The shape of evolution: a comparison of real and random clades. Paleobiology. 3:23–40.

Guyer C., Slowinski J.B. 1991. Comparisons of observed phylogenetic topologies with null expectations among three monophyletic lineages. Evolution. 45:340–350.

Harcourt-Brown K.G. 2002. Tree balance, time slices, and evolutionary turnover in Cretaceous planktonic foraminifera. Syst. Biol. 51:908–916.

Harcourt-Brown K.G., Pearson P.N., Wilkinson M. 2001. The imbalance of paleontological trees. Paleobiology. 27:188–204.

Hardman M., Hardman L.M. 2008. The relative importance of body size and paleoclimatic change as explanatory variables influencing lineage diversification rate: An evolutionary analysis of bullhead catfishes (Siluriformes : Ictaluridae). Syst. Biol. 57:116–130.

Harvey P.H., May R.M., Nee S. 1994. Phylogenies without fossils. Evolution. 48:523–529.

Heard S.B. 1996. Patterns in phylogenetic tree balance with variable and evolving speciation rates. Evolution. 50:2141–2148.

Heard S.B., Mooers A.O. 2002. Signatures of random and selective mass extinctions in phylogenetic tree balance. Syst. Biol. 51: 889–897.

Heath T.A., Zwickl D.J., Kim J., Hillis D.M. 2008. Taxon sampling affects inferences of macroevolutionary processes from phylogenetic trees. Syst. Biol. 57:160–166.

Jennings W.B., Pianka E.R., Donnellan S. 2003. Systematics of the lizard family pygopodidae with implications for the diversification of Australian temperate biotas. Syst. Biol. 52:757–780.

Jones K.E., Bininda-Emonds O.R.P., Gittleman J.L. 2005. Bats, clocks, and rocks: diversification patterns in chiroptera. Evolution. 59:2243–2255.

Kazancioglu E., Near T.J., Hanel R., Wainwright P.C. 2009. Influence of sexual selection and feeding functional morphology on diversification rate of parrotfishes (Scaridae). Proc R Soc B Biol. Sci. 276:3439–3446.

Kirkpatrick M., Slatkin M. 1993. Searching for evolutionary patterns in the shape of a phylogenetic tree. Evolution. 47:1171–1181.

Kozak K.H., Larson A.A., Bonett R.M., Harmon L.J. 2005. Phylogenetic analysis of ecomorphological divergence, community structure, and diversification rates in dusky salamanders (Plethodontidae: Desmognathus). Evolution. 59:2000–2016.

Kozak K.H., Weisrock D.W., Larson A. 2006. Rapid lineage accumulation in a non-adaptive radiation: phylogenetic analysis of diversification rates in eastern North American woodland salamanders (Plethodontidae : Plethodon). Proc. R. Soc. B Biol. Sci. 273: 539–546.

Li J., He Q.X., Hua X., Zhou J., Xu H.D., Chen J.K., Fu C.Z. 2009. Climate and history explain the species richness peak at mid-elevation for Schizothorax fishes (Cypriniformes: Cyprinidae) distributed in the Tibetan Plateau and its adjacent regions. Global Ecol. Biogeogr. 18:264–272.

Liow L.H., Quental T.B., Marshall C.R. 2010. When can decreasing diversification rates be detected with molecular phylogenies and the fossil record? Syst. Biol. 59:646–659.

Lloyd G.T., Davis K.E., Pisani D., Tarver J.E., Ruta M., Sakamoto M., Hone D.W.E., Jennings R., Benton M.J. 2008. Dinosaurs and the cretaceous terrestrial revolution. Proc. R. Soc. B Biol. Sci. 275:2483–2490.

Mann K.O., Lane H.R., editors. 1995. Graphic correlation. Tulsa (OK): SEPM Special Publication No. 53. Society Sedimentary Geology.

Marshall C.R. 1997. Confidence intervals on stratigraphic ranges with nonrandom distributions of fossil horizons. Paleobiology. 23:165–173.

McKenna D.D., Farrell B.D. 2006. Tropical forests are both evolutionary cradles and museums of leaf beetle diversity. Proc. Natl. Acad. Sci. U.S.A. 103:10947–10951.

McKenzie A., Steel M. 2000. Distributions of cherries for two models of trees. Math. Biosci. 164:81–92.

McLeish M.J., Chapman T.W., Schwarz M.P. 2007. Host-driven diversification of gall-inducing Acacia thrips and the aridification of Australia. BMC Biol. 5. doi:10.1186/1741-7007-5-3.

Merckx V., Chatrou L.W., Lemaire B., Sainge M.N., Huysmans S., Smets E.F. 2008. Diversification of myco-heterotrophic angiosperms: evidence from Burmanniaceae. BMC Evol. Biol. 8. doi:10.1186/1471-2148-8-178.

Mooers A.O. 1995. Tree balance and tree completeness. Evolution. 49:379–384.

Mooers A.O., Heard S.B. 1997. Evolutionary process from phylogenetic tree shape. Q. Rev. Biol. 72:31–54.

Moore B.R., Chan K.M.A., Donoghue M.J. 2004. Detecting diversification rate variation in supertrees. In: Bininda-Emonds O.R.P., editor. Phylogenetic supertrees: combining information to reveal the tree of life. Dordrecht: Kluwer Academic Publishers. p. 487–534.

Moore B.R., Donoghue M.J. 2007. Correlates of diversification in the plant clade Dipsacales: geographic movement and evolutionary innovations. Am. Nat. 170:S28–S55.

Moore B.R., Donoghue M.J. 2009. A Bayesian approach for evaluating the impact of historical events on rates of diversification. Proc. Natl. Acad. Sci. U.S.A. 106:4307–4312.

Near T.J., Benard M.F. 2004. Rapid allopatric speciation in logperch darters (Percidae: Percina). Evolution. 58:2798–2808.

Nee S. 2006. Birth-death models in macroevolution. Annu. Rev. Ecol. Evol. Syst. 37:1–17.

Paul C.R.C. 2009. The fidelity of the fossil record: the improbability of preservation. Palaeontology. 52:485–489.

Price T., Gibbs H.L., de Sousa L., Richman A.D. 1998. Different timing of the adaptive radiations of North American and Asian warblers. Proc. R. Soc. Lond. Ser. B Biol. Sci. 265:1969–1975.

Purvis A., Agapow P.M. 2002. Phylogeny imbalance: taxonomic level matters. Syst. Biol. 51:844–854.

Pybus O.G., Harvey P.H. 2000. Testing macro-evolutionary models using incomplete molecular phylogenies. Proc. R. Soc. Lond. Ser. B Biol. Sci. 267:2267–2272.

Quental T.B., Marshall C.R. 2010. Diversity dynamics: molecular phylogenies need the fossil record. Trends Ecol. Evol. 25:434–441.

Rabosky D.L. 2009. Heritability of extinction rates links diversification patterns in molecular phylogenies and fossils. Syst. Biol. 58:629–640.

Rabosky D.L. 2010a. Extinction rates should not be estimated from molecular phylogenies. Nature. 64:1816–1824.

Rabosky D.L. 2010b. Primary controls on species richness in higher taxa. Syst. Biol. 59:634–645.

Rabosky D.L., Donnellan S.C., Talaba A.L., Lovette I.J. 2007. Exceptional among-lineage variation in diversification rates during the radiation of Australia's most diverse vertebrate clade. Proc. R. Soc. B Biol. Sci. 274:2915–2923.

Rambaut A., Harvey P.H., Nee S. 1997. End-Epi: an application for inferring phylogenetic and population dynamical processes from molecular sequences. Comput. Appl. Biosci. 13:303–306.

Raup D.M. 1985. Mathematical models of cladogenesis. Paleobiology. 11:42–52.

Raup D.M., Gould S.J., Schopf T.J.M., Simberloff D.S. 1973. Stochastic models of phylogeny and the evolution of diversity. J. Geology. 81:525–542.

Ruber L., Van Tassell J.L., Zardoya R. 2003. Rapid speciation and ecological divergence in the American seven-spined gobies (Gobiidae, Gobiosomatini) inferred from a molecular phylogeny. Evolution. 57:1584–1598.

Ruta M., Pisani D., Lloyd G.T., Benton M.J. 2007. A supertree of Temnospondyli: cladogenetic patterns in the most species-rich group of early tetrapods. Proc. R. Soc. B Biol. Sci. 274:3087–3095.

Simpson G.G. 1944. Tempo and mode in evolution. New York: Columbia University Press.

Slowinski J.B., Guyer C. 1989. Testing the stochasticity of patterns of organismal diversity: an improved null model. Am. Nat. 134: 907.

Smith A.B., Pisani D., Mackenzie-Dodds J.A., Stockley B., Webster B.L., Littlewood T.J. 2006. Testing the molecular clock: molecular and paleontological estimates of divergence times in the echinoidea (Echinodermata). Mol. Biol. Evol. 23:1832–1851.

Steeman M.E., Hebsgaard M.B., Fordyce R.E., Ho S.Y.W., Rabosky D.L., Nielsen R., Rahbek C., Glenner H., Sorensen M.V., Willerslev E. 2009. Radiation of extant cetaceans driven by restructuring of the Oceans. Syst. Biol. 58:573–585.

Tolley K.A., Chase B.M., Forest F. 2008. Speciation and radiations track climate transitions since the Miocene climatic optimum: a case study of southern African chameleons. J. Biogeogr. 35: 1402–1414.

Vences M., Andreone F., Glaw F., Kosuch J., Meyer A., Schaefer H.C., Veith M. 2002. Exploring the potential of life-history key innovation: brook breeding in the radiation of the Malagasy treefrog genus Boophis. Mol. Ecol. 11:1453–1463.

Verdu M., Pausas J.G., Segarra-Moragues J.G., Ojeda F. 2007. Burning phylogenies: fire, molecular evolutionary rates, and diversification. Evolution. 61:2195–2204.

Wiens J.J. 2009. Paleontology, genomics, and combined-data phylogenetics: can molecular data improve phylogeny estimation for fossil taxa? Syst. Biol. 58:87–99.

Wiens J.J., Engstrom T.N., Chippindale P.T. 2006. Rapid diversification, incomplete isolation, and the "speciation clock" in North American salamanders (Genus plethodon): Testing the hybrid swarm hypothesis of rapid radiation. Evolution. 60:2585–2603.

Wiens J.J., Kuczynski C.A., Townsend T., Reeder T.W., Mulcahy D.G., Sites J.W. 2010. Combining phylogenomics and fossils in higher-level squamate reptile phylogeny: Molecular data change the placement of fossil taxa. Syst. Biol. 59:674–688.

Wiens J.J., Parra-Olea G., Garcia-Paris M., Wake D.B. 2007. Phylogenetic history underlies elevational biodiversity patterns in tropical salamanders. Proc. R. Soc. B Biol. Sci. 274:919–928.

Williams S.T., Duda T.F. 2008. Did tectonic activity stimulate Oligo-Miocene speciation in the Indo-West Pacific? Evolution. 62:1618–1634.

Yule G.U. 1924. A mathematical theory of evolution, based on the conclusions of Dr J. C. Willis. Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci. 213:21–87.