CONTENTS

6

SOME MODELS OF PHYLOGENETIC TREE SHAPE

Arne Ø. Mooers, Luke J. Harmon, Michael G.B. Blum, Dennis H.J. Wong and Stephen B. Heard

Abstract

As products of diversifying evolution, phylogenetic trees retain signatures of the evolutionary events and mechanisms that gave rise to them. Researchers have used a variety of theoretical models to represent different hypotheses about how diversification might proceed through the evolution of a clade. We outline two widely-used measures of phylogenetic tree shape, review a number of tree-generating models, and set out the predictions they make about tree shapes. The simplest of these models (the "Yule" and "Hey" models) are still used routinely, sometimes as if they provided good representations of diversification in Nature; but in fact, they do rather poorly when confronted with real data. More complex models that incorporate hypothesized macroevolutionary processes can in some cases provide a better fit to real data. We recommend further development of these more complex models - for instance, exploration of models that treat species as collections of individuals rather than as simple lineages. Much work remains to be done in estimating trees (especially waiting times), in exploring tree-generating models and in assessing patterns in the shapes of real phylogenies.

6.1 Introduction

Phylogenetic trees represent the evolutionary histories of lineages and so bear the impression of the evolutionary forces that gave rise to those lineages. Advances in molecular and computational techniques continually increase the number and size of our phylogenetic estimates. In the 1990's, both we [41] and Purvis [52] surveyed the two main aspects of phylogenetic tree pattern: variation in realized diversification rate among contemporaneous lineages, and changes in realized diversification rates through time. The techniques highlighted in these reviews have been used very successfully (see, e.g. [4, 10, 11, 62, 63]).

In parallel, researchers have continued to present generating process models for phylogenetic trees, in the hopes of being able to compare these with the real things. We offer a biological perspective on some of these models here. Our general thesis is that these models should do more than mimic reasonable tree

shapes: they should offer clear hypotheses that can be tested with the data as they become available. It is likely that real trees will be shaped by many factors and so these models should not be seen as mutually exclusive. All the models we survey are extensions of the simple birth-death process, in that evolving lineages have defined probabilities per unit time of giving birth to new lineages (causing a bifurcation) or terminating, and differ only in how these probabilities are assigned. We try to consider the strengths and problems of the models we survey and direct readers to some that we feel might show promise.

6.2   Background

We use the term tree shape to refer generically to both the distribution of sizes of the groups defined by nodes (called 'clades' by evolutionary biologists) and the distribution of edge weights (called 'branch lengths' by evolutionary biologists) on a directional bifurcating acyclic graph (Fig. 6.1). Our choice of graph structure is motivated by the fact that evolution is directional and primarily diversifying, and that events leading to multifurcations (i.e. vertices of degree > 3) are rare [29]. We recognize that our formulation overlooks other interesting graph structures relevant to evolution (e.g. cycles produced by recombination in gene trees or by hybrid species formation in species trees; uncertainty expressed in unrooted trees or in graphs with multifurcations). We further restrict ourselves to ultrametric trees, and refer to edge lengths and waiting times using time units. This is because we are interested in the actual diversification process through time, rather than in the inference process. This glosses over some painful facts - very few inferred trees have a robust timeline, and rooting trees is very difficult.

We concentrate on two aspects of tree shape. The first is the variation in subgroup size, captured very efficiently (see [2,36]) by Colless' measure of imbalance [9,23]. Colless' index I_c considers the number of leaves in the two partitions defined by each internal node (L and R) and is the sum of $|L - R|$ over all the $n - 1$ nodes in the tree, often normalized by the maximum possible value for a tree of size n (Fig. 6.1). Besides being the most commonly-used metric for bifurcating trees, it has a clear biological interpretation as an average measure of the realized differences in diversification rate of sister groups. Though $E(I_c)$ scales with n [23,58], its distribution has been characterized under the pure birth model [5]. I_c also has the property of being most sensitive to variation in clade size nearest the root of the tree [2,23].

The second aspect of tree shape is a measure of the distribution of nodes from the root to the leaves, or 'waiting times' (Fig.6.1), as captured in lineage-through-time plots [46,47]. Indices designed to summarize nodal distribution include stemminess [59] and γ [53] (or the closely related δ [54]). We return to these measures below.

As summary statistics, I_c and γ (or δ) do not capture all the variation within a sample of rooted trees [36]: for instance, two trees of size n with different topologies can nevertheless share the same I_c score [58]. They do, however, capture both differences in diversification rate among contemporaneous clades (I_c) and
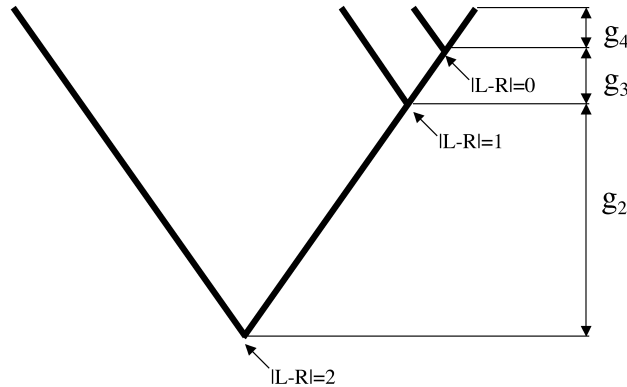
FIG. 6.1. A simple bifurcating tree highlighting the measures taken to summa-
    rize topology and waiting times. The sum of $|L - R|$ is used to give a measure
    of tree balance, while the waiting times g are used to create a measure of the
    relative placement of nodes between the root and the tips.

differences in diversification rate as one proceeds through time (γ). Importantly,
the two axes are not expected to vary independently [2, 41, 17]; unfortunately,
there is still little work that considers how reasonable diversification processes
affect both aspects simultaneously.

6.3   Yule and Hey models

The simplest possible model of diversification also has the oldest pedigree, devel-
oping from a simple model of diversification presented in 1924 [77]. Though the
original model had two parameters (one for the birth of species within genera
and one for the birth of genera), the "Yule Process" refers to a model where
there is no death, and diversification is modeled as a Markov process with a
single parameter λ, the instantaneous rate of birth [3]. The parameter λ can be
thought of as the average number of speciation events that occur in one lineage
per unit of time. The topologies of trees produced by this pure birth model can be
described recursively. At a node that subtends n species, the number of species
of the "left" subtree is chosen uniformly over $\{1, \ldots, n-1\}$ and this process is
continued in the left and right subtrees until the tips are reached [21, 70]. Blum
and colleagues [5] have recently derived the asymptotic distribution of I_c un-
der this model. The times during which there are i lineages, in the Yule model,
are exponential random variables with parameter λi [28, 47] and the conditional
probability distributions of the branching times given that n lineages are found
after t units of time can be found in [44] and [76]. Pybus and Harvey [53] took
advantage of constant λ to produce a standardized statistic γ such that $\gamma > 0$
with an increasing λ from the root to the tips, and $\gamma < 0$ with a slowdown in
diversification through time.

$$\gamma = \frac{\frac{1}{n-2}\sum_{i=2}^{n-1}\sum_{j=2}^{i} jg_j - \frac{T}{2}}{\frac{T}{\sqrt{12(n-2)}}},\qquad (6.1)$$

where T is given by

$$T = \sum_{j=2}^{n} jg_j.$$

The expression for γ is obtained after modifications of a test statistic introduced by Cox [12] in the context of Poisson processes (see Appendix).

The Yule model was extended to include a constant death rate (μ) by Raup and colleagues in the early 1970's [56]. Because the rates are invariant across lineages, this addition does not change the expected distribution of topologies. However, because we will now sample lineages in the present time that are destined to go extinct [22], there are "too many" lineages near the present, and $\gamma > 0$.

Another simple model for generating tree shapes was presented by Hey [28]. This model assumes that the total number of species N is fixed and that each lineage bifurcates at rate λN. More precisely, the time before speciation in each lineage is an exponential random variable with parameter λN. Furthermore, at each speciation event a lineage chosen uniformly among all lineages goes extinct, insuring that the total number of species remains constant. The same model is known in population genetics as the Moran model [42] [13, page 18-23]. What is usually called the Hey model, though, is not the forward-in-time model that is described above but the backward-in-time model that corresponds to the genealogy of n species sampled among the N extant species. Therefore it should be emphasized that Yule trees describe the genealogies of entire monophyletic groups whereas Hey trees describe, within a monophyletic group, the genealogies of samples of species ($n \leq N$). The genealogy in the Hey model is equivalent to a well-characterized model known in population genetics as the coalescent [27] [Felsenstein, this volume]. The topology of the Hey (coalescent) process is simply described as follows: starting with n lineages, two pairs of lineages are chosen uniformly among all the possible pairs to coalesce and this coalescence process is continued until there is only one remaining lineage. The topologies of the Hey trees (and so its measure, I_c) are distributed identically to the topologies of the Yule trees [47] . In the Hey model, the time during which there are exactly i lineages is an exponential random variable with parameter $\lambda i(i-1)$ [28, 47]. Note that the expected values of coalescence times $1/i(i-1)$ (when $\lambda = 1$ as it is usually assumed in the Moran model) in the Hey model differ by a factor 2 to the expected values of coalescence times $2/i(i-1)$ in the coalescent model as it is usually used in population genetics [13, page 23]. Under the Hey model, the statistic γ is expected to be large, with more nodes found nearer the leaves than under the Yule model. Pybus and colleagues [54] took advantage of the known

waiting times for events under the coalescent to produce a new standardized measure denoted δ

$$\delta = \frac{\frac{T'}{2} - \frac{1}{n-2} \sum_{i=n}^{3} \sum_{j=n}^{i} j(j-1)g_j}{\frac{T'}{\sqrt{12(n-2)}}}, \tag{6.2}$$

where T' is given by

$$T' = \sum_{j=n}^{2} j(j-1)g_j.$$

The expression of δ given by Pybus et al. [54, equation (6.2)] results from our equation (6.2) after dividing the numerator and the denominator of our equation (6.2) by 2. The derivation of the statistic δ is given in the Appendix. We note that Pybus did not apply δ to species trees.

Importantly, both these models do a remarkably poor job of capturing the distribution of tree shapes reported in the literature [41, 2, 6, 30, 69]: published trees are much more imbalanced (have higher I_c values) than expected. This is an important and perhaps still underappreciated finding: if our published trees are unbiased with respect to shape, there are strong macroevolutionary forces at work that demand explanation. However, perhaps because of their convenience, these null models are still often used either explicitly [45, 44, 78] or implicitly (see, e.g. [7]).

6.4 $\lambda = \textbf{function(trait)}$

The core assumption of the models presented above, that all species have equal speciation rates at a given time, is an assumption that most evolutionary ecologists would always have rejected. Instead, at least since Darwin's time, an enormous amount of attention has been paid to the notion that some lineages might experience higher speciation rates (or lower extinction rates) than others, either due to intrinsic properties of the species, extrinsic factors having to do with the environment, or the interaction of the two [43, 25, 62]. Differences in diversification rates among related lineages have in fact been documented for a variety of clades (e.g. [7, 38, 67]), and analyses of branch-length distributions in phylogenies [61] have established that differences in diversification rate not only exist, but tend to be propagated along evolving lineages (such that high or low rates are 'heritable' from ancestral to descendent species). An important class of generating models [24] seeks to incorporate some of this biology by considering the case where the speciation rate λ is a (perhaps nonlinear) function of some variable x, where x takes on a value for each species that is determined by an evolutionary model over the phylogeny of an evolving clade. Most simply, x can be interpreted as any evolving trait (simple or complex) of the organisms, such as body size, dispersal rate, feeding strategy or pollination syndrome [24], but it could equally

represent a characteristic of the environment, so long as restricted dispersal by the organisms constrained the value of x for one species to resemble the value of x for its ancestor. In either case, λ varies among species in an evolving clade, but does so with nonzero heritability (there is a resemblance between ancestor and descendent) such that whole lineages are typified by higher or lower speciation rates.

Heard [24] explored a model belonging to this class, in which a trait value x evolved in a clade by a random walk, with changes either gradual (continuous in time) or punctuated (occurring only at speciation events). In this model, λ for each species was a simple function of the trait value x, plus a 'noise' term ϵ representing other influences on speciation rate. Heard [24] found that this model produced phylogenies with high I_c compared to the ERM, and that I_c values typical of real phylogenies could be produced—albeit with high rates of evolution in the trait value x (or, more generally, in the rate of evolution of the diversification rate parameter itself). Furthermore, speciation-rate variation arising through the addition of the 'noise' term increased I_c, but only when values of ϵ were persistent through time (that is, when ϵ changed only at speciation events, rather than continuously through time). This model drew attention to the importance of differences in diversification rates that are maintained by lineages through time (either through trait heritability or through other temporally persistent effects on λ) in generating phylogenies with high I_c. Efforts to demonstrate the existence of heritable diversification-rate variation [61] and to devise tests for correlates of diversification rate (see, e.g. [50]) were inspired directly by this generating model.

Heard [24] did not consider the nodal height distribution property of the trees produced by his model. Because clades in Heard's model become dominated by high-diversification-rate lineages [24] via species sorting [74], we would expect their phylogenies to have $\gamma > 0$ as more speciation events occur closer to the present. However, whether models of this type can produce trees with realistic values of γ (and do so for the same parameter values that produce realistic I_c) remains unknown.

6.5 $\lambda =$ **function(age)**

In this class of generating models, λ varies among species only as a function of the time elapsed since a species's last speciation event (its age). One can imagine biological circumstances under which speciation rates might be either higher or lower for young species, and both cases have been modeled.

Models in which young species have smaller λ are biologically plausible when young species tend to have small population sizes or small geographic ranges. This is, in fact, a prediction of most models of speciation, most notably of the peripheral-isolate model [39]. Two slightly different models have been proposed. Losos and Adler [35] described a model in which speciation rate $\lambda = c$ for all lineages, except that following speciation, one daughter lineage has $\lambda = 0$ during a refractory period of length a^*. As an alternative, Chan and Moore [8] modeled λ

as increasing linearly from zero to c over a period a^* for both daughters following a split. In either case, with a^* small to moderate compared to total tree height, these models produce phylogenies more balanced (lower I_c) than does the pure-birth model. (When a^* is a substantial fraction of total tree height, the resulting phylogenies have higher I_c than pure-birth, but such large values of a^* are probably not plausible in the biological context that inspired the models). Because these models, then, produce phylogenies even more unrealistic than the pure-birth model ("real" trees have higher I_c than pure-birth, not lower), they have not attracted much recent attention. Our preliminary work (SBH and DHJW) suggests that reasonable values of a^* have no effect on γ. Moderate refractory periods lower the effective speciation rate, but do not change the relative distribution of speciation events over the height of the tree. Much longer refractory periods do give rise to trees with negative γ, but again, such long a^* are probably unrealistic.

Models in which young species have larger λ are biologically plausible when speciation events are likely to occur in bursts—for instance, because lineages that are speciating have colonized a new region, and a new region with many open niches favours multiple speciation events [70]. Agapow and Purvis [2] considered a discrete time model in which λ increases after speciation, followed by decay back to $c : \lambda(a) = c + Ka^{-0.5}$, where a is age (time post speciation, with both daughters of a speciation event beginning with age $a = 0$). Steel and McKenzie [70] examined a general class of models in which λ decreases monotonically with a (the Agapow and Purvis model is a special case), but developed in particular a subclass in which $\lambda(a) = 0$ for $a > m$, where m is a constant speciation window. A simple version of this model, essentially the converse of the Losos-Adler refractory period model, would have $\lambda(a) = c$ for $a \leq m$, and $\lambda(a) = 0$ for $a > m$. Both the Agapow-Purvis and the Steel-McKenzie models produce imbalanced phylogenies (high I_c, which is realistic), and distributions of nodal heights with more speciation events towards the root of the tree ($\gamma < 0$). However these models generally have been explored only by simulation; formal results establishing distributions of I_c or γ are known for only a few special cases (see, e.g. [5]).

There are (at least) two interesting questions one could ask about Agapow-Purvis Steel-McKenzie models. The first of these is statistical, and concerns the ability of the models to produce trees with any given distribution of shapes. The second question is more biological, and concerns the fit of model results to real-world trees.

The Steel-McKenzie model was motivated by the Uniform distribution of phylogenies, a natural distribution of interest to many mathematicians whereby all labeled cladograms (rooted trees where the branch lengths are not considered) are equally likely. Under this distribution, trees are random guesses [68]. This model might be useful as a prior for Bayesian tree inference. However, despite its mathematical attractiveness, evolutionary biologists have largely failed to imagine plausible process models that produce such a distribution. Steel and

McKenzie [70] proved that their model does produce the Uniform distribution when $\lambda(a) = 0$ for $a > m, m < \frac{T}{n}$, where T is an arbitrary time horizon. However, upon closer examination this result appears to be of primarily mathematical interest because the trees produced under these conditions are not biologically plausible. To see this, consider that any lineage that fails to speciate before a time m since its birth is a spinster that can never speciate again; and a tree in which all lineages are spinster lineages is a spinster clade that can never increase in size. But the condition for producing the Uniform distribution is $m < \frac{T}{n}$, or $T > mn$. Since each lineage must speciate within an interval m or become a spinster, after a period $T > mn$ the only trees of n lineages that can exist are spinster trees. We do not believe that many (if any) real clades are spinster clades in which the origin of new lineages is no longer possible; on the contrary, available evidence suggests that speciation continues today in many if not all clades (e.g. [62, 71]).

This does not mean, however, that the model should be discarded. Instead, one can ask the second question about the model: can it produce realistic tree shapes with plausible parameter values? Using the same approach we have described for other model classes, one could compare Steel-McKenzie model trees with collections of real estimated phylogenies, asking whether plausible values of model parameters (c, K, m, or others in more complex models of the class) can produce phylogenies with realistic I_c and γ. This is an open question, in part because what constitutes "realistic" γ values is not well established, and in part because the biological or palaeontological data needed to assess plausibility of a particular choice of K or a^* are not obviously available. Analysis of this sort (as in [24]) is logically straightforward, at least, and could establish whether 'speciation-burst' biology is a good candidate as a contributor to the shapes of real trees.

6.6 $\lambda = \mathbf{function(time)}$

There are several verbal models that make λ a declining function of absolute time rather than the age of the lineage; for instance, key innovations or new biogeographic opportunities may allow for an initial flourish of speciation that then settles down. However, the model that has received the most attention is that of adaptive radiation (AR [62]). Adaptive radiation is the evolution of phenotypic divergence in a rapidly multiplying lineage [62]; indeed, it is primarily the emphasis on phenotypic divergence that separates AR models from the models considered in the previous section. Some claim that adaptive radiation may account for much or even most of present day diversity (D. Schluter, pers. comm.). One expectation from AR theory is that speciation is rapid in its initial stages and then slows down (so, e.g., $\gamma < 0$; [62, 19]). This seems to be the case for some fossil [18] and some extant clades [46, 51, 60, 66]. One presumed underlying pattern has clades growing rapidly and then, as birth rates decline below extinction rates, shrinking. We note that this particular trajectory has been formally modeled for species numbers by Raup and colleagues [56] and Strathman and

Slatkin [72] and presented as an example for waiting times on trees by Nee and colleagues [47].

More quantitative work on AR tree shape is needed. Gavrilets and Vose [19] have made a start with an individual-based simulation approach to AR, where sexual diploid individuals with complex genomes evolve on discrete patches arranged on an initially empty but heterogeneous grid. These individuals migrate, undergo selection and eventually form populations that speciate. They found that speciation was vastly more common during the early stages of the diversification; resulting trees would have low γ values. They also often observed 'overshooting', where the clade size at the end of a run was smaller than the maximum reached during a run. Though they do not look at tree balance, Gavrilets and Vose [19] interpret some of their simulation results in light of a verbal model of a few generalist lineages rapidly speciating into slower-evolving specialists, which might give rise to imbalanced trees. The generalist to specialist pattern is, however, not strongly supported by available comparative data [49, 62].

6.7   The Neutral model

Another rich if controversial approach to explaining biodiversity production is Hubbell's "Unified Neutral Theory" or UNT [31]. The UNT has at its core a metacommunity landscape saturated with competitively identical individuals. This landscape is made up of patches that can be occupied by only one individual, regardless of its species. In this model, individuals in the metacommunity compete for space, with patches vacated by death filled by migration of a new individual from surrounding patches. This feature makes the UNT a null model for community organization and evolution, and it is widely agreed that at least some communities deviate strongly from the UNT. However, the extent to which this is true is currently under intensive debate (e.g., [20, 40]). Although much of the focus of this debate has been placed on the ability of this theory to explain relative abundances within communities (e.g. [16, 73]), the UNT also makes predictions about the shapes of phylogenetic trees. In the context of diversification, Hubbell's model has an unchanging per individual speciation rate over the entire metacommunity, while extinction occurs whenever the population size of any species reaches zero individuals. As a consequence, per-species speciation and extinction rates are a function of population size. Critically, under the UNT extant lineages differ in a predictable fashion in relative abundance, collectively approximating a truncated log-normal distribution. This means that at any time, extant lineages differ predictably in their propensities to speciate and to go extinct. Hubbell [31] was able to demonstrate by simulation that the UNT produces trees with a concentration of short branches near the tips (high γ values), because extinction is highest when there are many species at small population size.

The UNT is qualitatively different from AR in that lineages do not evolve to take advantage of heterogeneous resources. Also, because speciation is conceived of as a point mutation in one individual, its behaviour in terms of population

size is punctuational [15]—the parent species is very similar in size before and after the speciation event while the daughter lineage is made up of a single individual and so it initially has a very low probability of speciating and a very high probability of going extinct.

In order to address what sorts of trees this explicitly ecological model produces, we modeled the UNT for a metacommunity composed of 441 local communities arranged in a 21×21 grid. Each local community was made up of 100 individuals for a total metacommunity size (J_m) of 44,100. Hubbell [31] defines a "fundamental biodiversity number" $\theta = 4*J_m*v$, where v is the per-capita speciation rate. For our simulations, we used a value of $\theta = 10$. Following Hubbell [31], we ran simulations in discrete time and allowed one individual per local community to die and be replaced by a birth, migration, or speciation event in each generation. We limited migration to communities that were immediate neighbors in the grid [31]. We then simulated community drift and diversification under a range of migration rates. We started each simulation with a metacommunity completely filled with individuals of a single species, and ran them until both species abundance distributions and phylogenetic tree shape reached a dynamic equilibrium. For this set of parameter values, tree shape equilibrium was reached at around 100,000 generations, but to ensure that our results represent tree shapes at metacommunity equilibrium, we ran simulations for 500,000 generations to produce phylogenetic trees. Increasing migration rates had a negative impact on metacommunity species richness (Fig.6.2; [31]). As stated by Hubbell [31], we found that phylogenetic trees generated from these simulations show a concentration of short branches near the tips; as a consequence, γ values were consistently high over a range of migration rates. In fact, for most sets of simulations, over half of the produced phylogenies had $\gamma > 2$, and would constitute a significant deviation from the pure birth expectation. This effect was most pronounced for very low migration rates, but that may be influenced by higher power of the γ statistic [53] for the larger trees such simulations produce. Phylogenies produced by this model were highly imbalanced; in fact, most phylogenies were completely pectinate trees (Fig.6.2). The percentage of completely pectinate trees increased with higher migration rates. This is because under Hubbell's model, metacommunities with high migration rates have a steeper rank abundance curve [31]. Since variation in speciation rate in a metacommunity is related to the slope of the rank abundance distribution, communities dominated by a single abundant species will have more imbalanced phylogenetic trees than communities with a more even abundance distribution. This prediction is probably robust to many aspects of Hubbell's model, and follows from the mode of speciation and relationship of speciation rates to abundances.

Formal comparisons of UNT tree-shape predictions with the shapes of real phylogenies have not been conducted, but it seems fairly clear that the UNT as we implement it above produces trees that are much too imbalanced to be realistic (it is more difficult to assess predictions for γ, since the distribution of γ for real trees is not known). This is an interesting result, since other modeling
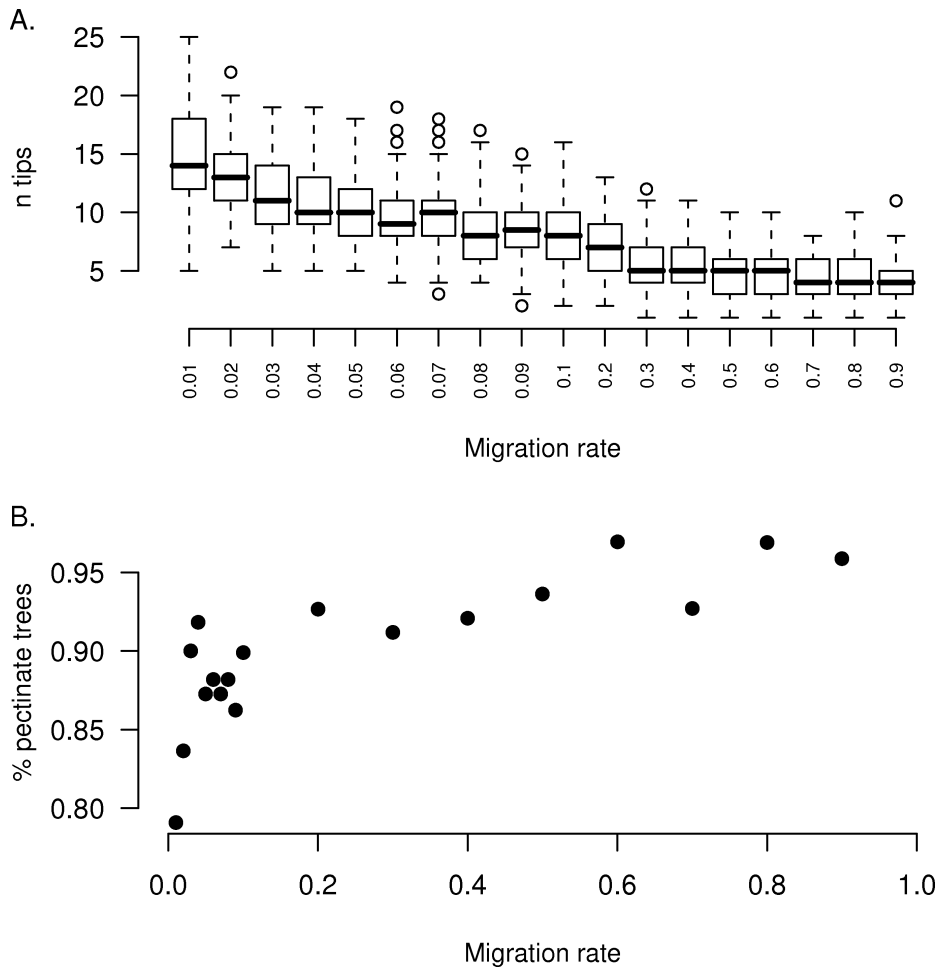
A.

B.

FIG. 6.2. The behaviour of diversification under the Hubbell's Unified Neutral
   Model. (A) the average size of trees with increasing migration rate among
   patches in the metacommunity. (B) The proportion of fully pectinate trees
   at equilibrium for communities with different migration rates among patches
   in the metacommunity. Because many of the trees at high migration rates
   are small, this is a better measure of tree shape than standardized I_c. For all
   runs, $J_m = 44,100$ and $\theta = 10$.

efforts have found it difficult to produce trees that are imbalanced enough for
realism [24]. It is unknown whether elements of the UNT assumptions (such
as population-size dependence), might in a more sophisticated model be able
to produce realistic distributions of I_c and γ; this is an area for further study.
Hubbell [31] shows, in addition, that species abundance distributions are much

more even under a "fission" model of speciation, where speciation events involve randomly dividing the ancestral population into two parts; we predict that this mode of speciation will result in more balanced phylogenetic trees than those produced in our implementation (Fig. 6.2). However, there will still be some relationship between ancestral and daughter population sizes, and trees will likely be more imbalanced than those produced under the Yule model.

6.8 $\quad \lambda = \text{function}(N)$

One feature missing from all models discussed so far is any tendency for diversity to be limited—that is, for diversity to reach an equilibrium N^* analogous to carrying capacity in the logistic model of population growth. Such an equilibrium will result if per-capita extinction rates increase, or per-capita speciation rates decrease, with standing diversity. Such effects are plausible for a variety of biological reasons—for instance, if high diversity means smaller population sizes for each species, raising extinction risk. However, whether such limits to diversity are ever reached in Nature is an open question. Paleontologists have modeled diversity in the marine fossil record with logistic-like functions that assume limits to diversity, with some success for Paleozoic faunas but more debatable results for Mesozoic and Quaternary faunas [64,65]. Ecologists have also devoted considerable theoretical and empirical attention to the idea of "limiting similarity" in communities (and by extension, clades), which would impose limits to diversity by setting a maximum number of niches available for occupation [1, 32, 34, 37]. A half-century of research, though, has produced no consensus on whether such models explain much about real communities. Indeed, some models of diversification either assume or imply that diversification is more likely to proceed with positive feedback than with negative: for instance, escape-and-radiation [14] and cascading host-race formation [71].

Surprisingly, little is known about tree shapes under models of limited diversification. Harvey et al. [22] considered a model in which extinction rate increases with diversity, but speciation rate is constant. However, they did not report balance for their model, and (considering only the extant species) only report that nodal height distributions are similar to those from a mass-extinction model. More complex models, with both speciation and extinction rates responding to diversity, show more complex behaviour (DHJW and SBH, unpublished data), for instance with γ depending strongly on the ratio of speciation to extinction rates at half of N^*. Few studies have yet asked whether limited-diversity models produce tree shapes typical of real clades, although Nee et al. [46] interpreted the shape of a compound bird phylogeny as consistent with niche-filling model (though one with diversification rate decreasing to zero only as N^* approaches infinity).

A rather different approach to modeling limited diversity is implicit in the simple Hey model [28]. In Hey's model, a clade reaches size N^* and subsequently each speciation event (as speciation continues with constant rate) is balanced by a randomly imposed extinction event. Notably, the Hey model is mute with
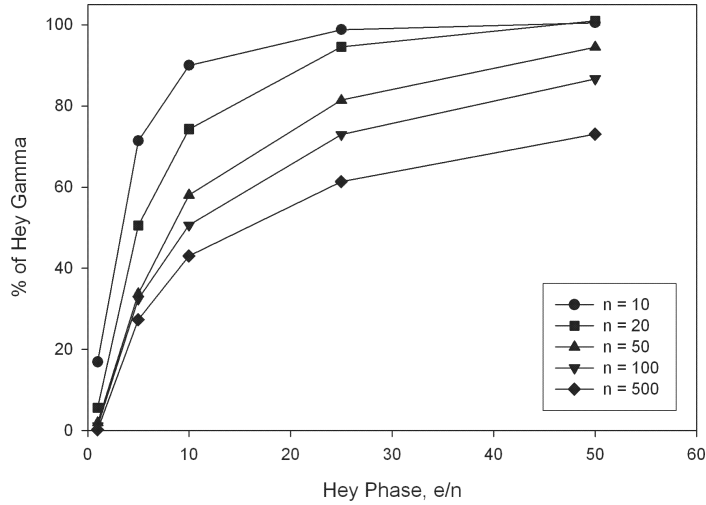
FIG. 6.3. Approach to stationary γ distribution for trees grown to size n under a Yule model, followed by balanced speciation and extinction under Hey's [28] model. The length of the Hey phase is measured as number of speciation/extinction events (e) as a multiple of the number of species in the tree (n), and γ expressed as a percentage of that expected under the Hey model.

regard to how a clade reaches size N^* [47]. So, for instance, Zhaxybayeva and Gogarten [78], who recently used the model to simulate the early tree of life, simply start with N^* unrelated lineages and allow the model to run until all the extant individuals have a single common ancestor (all other $N^* - 1$ lineages having died out). Another approach that better mimics radiations is to consider a two-phase process: a tree first grows to size N^* ("growth phase"), followed by some time spent at size N^* ("Hey phase"). So long as the tree's Hey phase is long enough to reach the stationary distribution of tree shapes (that is, for any signature of the growth phase to be erased), the growth phase model doesn't matter. But how long a Hey phase might be required to reach the stationary distribution, and is this plausible for real trees?

This question has not been addressed for any growth phase model, but we can make a start by examining one simple possibility: growth phase diversification under the Yule model. We implemented a simulation model (following [23, 24]) of tree growth under the Yule model, followed by speciation and extinction (still at a constant rate for all lineages) in a Hey phase of variable length. We measure the length of a Hey phase in terms of species turnover: if there are N^* species when the Hey phase begins, then a Hey phase of length 1 has N^* speciation (and N^* extinction) events; the average species is replaced in the phylogeny once. We generated 500 replicate trees of $N^* = 10, 20, 50, 100$ and 500, with Hey phases of length 1, 5, 10, 25, and 50. We consider a Hey phase of even

length 10 to be extraordinarily long, as it implies that since the clade reached its equilibrium diversity N^*, each species has (on average) been replaced 10 times over; or alternatively, over 90% of the history of the clade has been spent at equilibrium diversity. Since our Yule trees start with the same distribution of I_c as expected following the Hey phase [47], there is no change in this attribute of tree shape (as there might be under other growth-phase generating models). The nodal height distribution does, however, change: the Yule trees that enter the Hey phase have growth-phase $\gamma = 0$ [53], (we confirm this in our simulations), whereas Hey trees will have large, positive γ. Importantly, for trees of moderate to large size, the approach to stationary Hey-phase γ is quite slow (Fig. 6.3): for instance, a Hey phase of length 10 brings trees of $n = 50$ and $n = 500$ trees just 58% and 43% respectively of the way from the growth-phase γ to stationary Hey-phase γ.

Since we have little evidence that modern clades are at an equilibrium diversity (N^*) at all, let alone that clades spend much time at N^*, we conclude that the Hey model is probably not very relevant to the shapes of real phylogenies. Of course, our use of the Yule model for the growth phase can (and should) be criticized, but we do not expect this changes the overall picture much. Indeed, because the Hey model produces trees with the Yule distribution of topologies, it does not mimic the trees we infer from Nature.

6.9   Concluding Remarks

Since the 1980's, we have known that the simplest models do a poor job of modeling the shapes of published phylogenetic trees. Tree reconstruction methods may be biased with respect to shape [41], but current surveys suggest the problem may not be grave for trees $<\sim 50$ tips [69], which is good news. However, Wilkinson et al. [75] point to some clear biases with supertree construction methods, and now that we are routinely reconstructing very large trees ($N \gg 100$ tips) using heuristic methods, we continue to urge caution. A related problem is taxon choice: sampling methods that inadvertently lead to unbalanced trees are easy to conceive of, but little empirical or theoretical work has been done in this area. However, our intuition is that the tree of life is very imbalanced at all levels. There may be different processes occurring at different temporal scales; if so, we must be careful what we infer from our biologically-motivated models. Even at any one scale, different processes likely co-occur in Nature, and at different times during the history of a lineage. Also, the same model applied at different scales can produce different patterns. For instance, the last two sections above consider two models (UNT and Hey) that are each based on the coalescent. However, the expected shapes are very different because the first applies the coalescent to individuals that are then aggregated into species, and the second applies the coalescent to the species themselves. This scale issue may be interesting to pursue.

We know of no general survey of waiting times (e.g. looking at γ or δ) for ultrametric trees; Hey [28] considered eight small trees but could reject neither

the Yule nor the Hey models for any, suggesting very low power. Getting branch lengths that are proportional to time is very difficult (see, e.g. [60]), and advances in this area are needed. However, from the modeling perspective, there is room for much more work. As an example, the Adaptive Radiation and Hey models both have aspects of diversity-dependent cladogenesis [56], but they make strongly contrasting predictions about waiting time: AR models should produce many longer terminal branches than the Hey model.

Our approach in this chapter has been to consider simple tree measures as reflections of the underlying macroevolutionary models and to compare these to expectations under various null models. A more powerful approach would use maximum likelihood and information criteria to select among alternative models directly. Building on the pioneering work on birth-death models by Nee and colleagues [47], such tools have been created [55], but more work is needed to confront a wider range of possible models of diversification [54].

In the introduction, we stated that tree shape is the record of the tempo and mode of diversification. For evolutionary biologists, being able to explain these patterns is a major intellectual goal. However, we would like to end this chapter by pointing out that the study of tree shape resonates at a more immediate level. Rauch and Bar-Yam [57] point to the fact that reasonable models of gene genealogies imply that genetic diversity within species is highly skewed. Non-random loss of the few individuals bearing divergent genes can greatly decrease a species's overall genetic diversity. The same holds true at larger scales: to the extent that phylogenies are highly imbalanced and have long terminal branches, a few lineages will represent much of the tree's total diversity and their extinctions would represent disproportionate loss to the tree of life [26][Hartmann and Steel, this volume]. Given the rate at which we are losing species through anthropogenic extinctions, understanding this may help us in efforts to preserve the products of evolution for the distant future.

6.10   Acknowledgements

6.11   Appendix

The statistics γ and δ that have been introduced by Zink and Slowinski [79] and Pybus and colleagues [53, 54] in order to detect trends in speciation rates have been derived from a test statistic that can be found in [12]. In [12], the null model is a simple homogeneous Poisson process and the alternative hypothesis
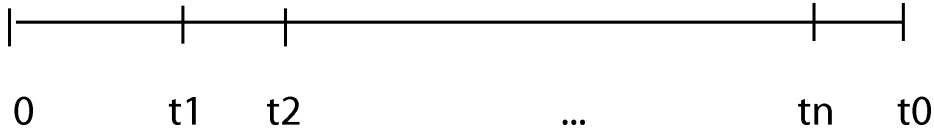
0   t1   t2   ...   tn   t0

Fig. 6.4. An illustration of a Poisson process. The $t_i's$ denote the time at which events occurred. In a homogeneous Poisson process, the $(t_{i+1} - t_i)$'s are exponentially distributed with the same parameter λ.

corresponds to a model where the instantaneous rate of occurrence λ is not constant anymore but varies with time. In a homogeneous Poisson Process, the times between successive events are exponentially distributed with the same parameter λ.

An important result concerning homogeneous Poisson processes is that the joint probability distribution of $(t_1, \ldots, t_n)$ conditional on observing n events between time 0 and time t_0 is the same as the joint probability distribution of an ordered vector of independent and uniform (over $(0, t_0)$) random variables (Fig. 6.4).

Thus, because the value of the sum of the t_i's is invariant by permutation, we have

$$E[\sum_{i=1}^{n} t_i] = n\frac{t_0}{2}. \tag{6.3}$$

Here we used the fact that the expected value of a uniform random variable over $(0, t_0)$ is $t_0/2$. Using equation (6.3) and the fact that $\text{Var}[\sum t_i] = nt_0^2/12$, Cox [12] introduced the following test statistic

$$S = \frac{\sum_{i=1}^{n} t_i - n\frac{t_0}{2}}{t_0\sqrt{\frac{n}{12}}}. \tag{6.4}$$

This statistic captures the fact that the mean of the t_i's should be around $t_0/2$ if there is no trend (i.e. λ = constant), larger than $t_0/2$ if the rate increases with time ($t_i's$ will be nearer t_0) and smaller than $t_0/2$ if the rate decreases with time ($t_i's$ nearer to 0) .

Sometimes, events are not observed until a given point of time but until, let's say, the nth event. In that case $t_n = t_0$ and n should be replaced by $n - 1$ in equation (6.4) (see [12])

$$S = \frac{\sum_{i=1}^{n-1} t_i - (n-1)\frac{t_0}{2}}{t_0\sqrt{\frac{n-1}{12}}}. \tag{6.5}$$

In the Yule process, the inter-speciation times are still exponentially distributed but the parameter of the exponential random variables varies with the

number of species. If we denote by g_i's the inter-speciation time corresponding to the time during which there are exactly i species (see Fig. 6.1), the random variable g_i is exponentially distributed with parameter λi.

In order to build a test statistic, the trick consists of considering the ig_i's rather than the g_i's because the ig_i's are exponentially distributed with the same parameter λ. Equation (6.4) can then be used by first noting that the first event (at the root) gives no information (i.e. it only defines $t = 0$). We therefore have only $n - 2$ observations. The test statistic is then obtained from equation (6.4) after replacing n by $n - 2$, t_i by $\sum_{j=2}^{i+1} jg_j$, t_0 by $\sum_{j=2}^{n} jg_j$ and simplifying the summands in the first term

$$\gamma = \frac{\sum_{i=2}^{n-1}\sum_{j=2}^{i} jg_j - \dfrac{n-2}{2}\sum_{j=2}^{n} jg_j}{\sum_{j=2}^{n} jg_j \sqrt{\dfrac{n-2}{12}}}. \tag{6.6}$$

In the Hey process, speciation events should be viewed backwards. The g_i's are exponentially distributed with parameter $i(i-1)\lambda$ and the root corresponds to the last (viewed backward) event. Because the end of the time interval during which the process is observed corresponds to a speciation event (the first one at the root), equation (6.5) should be used instead of equation (6.4) and the test statistic becomes

$$\delta = -\frac{\sum_{i=n}^{3}\sum_{j=n}^{i} j(j-1)g_j - \dfrac{n-2}{2}\sum_{j=n}^{2} j(j-1)g_j}{\sum_{j=n}^{2} j(j-1)g_j \sqrt{\dfrac{n-2}{12}}}. \tag{6.7}$$

The minus sign is introduced so that that the statistic is positive when nodes are closer to the tips than expected and negative conversely. The reason why the sum ranges from n to 2 or 3 is that the speciation process should be viewed backwards in the Hey (coalescent) process. Thus, the "first" speciation event occured g_n units of time before the present, the "second" speciation event occured g_{n-1} units of time before the "first" speciation event, and so on. The statistics introduced by Pybus [53, 54] (our equations (6.1) and (6.2)) are then given by equation (6.6) and (6.7) after dividing their numerators and their denominators by $n - 2$.

REFERENCES

[1] Abrams, P.A. (1983). The theory of limiting similarity. *Annual Review of Ecology and Systematics*, **14**, 359-376.

[2] Agapow, P.M. and Purvis, A. (2002). Power of eight tree shape statistics to detect nonrandom diversification: A comparison by simulation of two models of cladogenesis. *Systematic Biology*, **51**, 866-872.

[3] Aldous, D.J. (2001). Stochastic models and descriptive statistics for phylogenetic trees, from Yule to today. *Statistical Science* **16**, 16-34.

[4] Baldwin B.G and Sanderson M.J. (1998) Age and rate of diversification of the Hawaiian silversword alliance (Compositae) *Proceedings Of The National Academy Of Sciences Of The United States Of America* **95**: 9402-9406.

[5] Blum, M.G.B., Francois, O., and Janson, S. (2006). The mean, variance and joint distribution of two statistics sensitive to phylogenetic tree balance. *Annals of Applied Probability*, In Press.

[6] Blum, M.G.B. and Francois, O. (2006). Which random processes describe the Tree of Life? A large-scale study of phylogenetic tree imbalance. *Systematic Biology*, **55**, 685-691.

[7] Cardillo, M., Orme, C.D.L., and Owens, I.P.F. (2005). Testing for latitudinal bias in diversification rates: an example using New World birds. *Ecology*, **86**, 2278-2287.

[8] Chan, K.M.A. and Moore, B.R. (1999). Accounting for mode of speciation increases power and realism of tests of phylogenetic asymmetry. *American Naturalist*, **153**, 332-346.

[9] Colless, D.H. (1982). Phylogenetics: the theory and practise of phylogenetic systematics II (book review). *Systematic Zoology*, **31**, 100-104.

[10] Cotton, J.A. and Page, R.D.M. (2005). Rates and patterns of gene duplication and loss in the human genome. *Proceedings of the Royal Society of London Series B—Biological Sciences*, **272**, 277-283.

[11] Cotton, J.A. and Page, R.D.M. (2006). The shape of human gene family phylogenies. *BMC Evolutionary Biology*, **6**, 66.

[12] Cox, D.R. and Lewis, P.A.W.L. (1966). *The Statistical Analysis of a Series of Events*. Chapman and Hall.

[13] Durrett, R. (2002). *Probability models for DNA sequence evolution.* Springer-Verlag.

[14] Ehrlich, P.R. and Raven, P.H. (1964). Butterflies and plants: a study in coevolution. *Evolution*, **18**, 586-608.

[15] Eldredge, N. and Gould, S.J. (1972). Punctuated equilibira: an alternative to phyletic gradualism. In *Models in Paleobiology* (ed. T. J. M. Schopf and J. M. Thomas). pp. 305-322, Freeman, Cooper, San Francisco.

[16] Etienne, R. S., and Olff, H. (2005). Confronting different models of community structure to species-abundance data: a Bayesian model comparison. *Ecology Letters*, **8**, 493-504.

[17] Felsenstein, J. (2004). *Inferring Phylogenies*. Sinauer Associates, Sunderland, Massachusetts.

[18] Foote, M. (1993). Discordance and concordance between morphological and taxonomic disparity. *Paleobiology*, **19**, 185-204.

[19] Gavrilets, S. and Vose, A. (2005). Dynamics of adaptive radiation. *Proceedings of the National Academy of Sciences (USA)*, **102**, 18040-18045.

[20] Gilbert, B. and Lechowicz, M.J. (2004). Neutrality, niches, and dispersal in a temperate forest understory. *Proceedings of the National Academy of Science (USA)*, **101**, 7651-7656.

[21] Harding, E. (1971). The probabilities of rooted tree-shapes generated by random bifurcation. *Advances in Applied Probability*, **3**, 44-77.

[22] Harvey, P.H., May, R.M., and Nee, S. (1994). Phylogenies without fossils. *Evolution*, **48**: 523-529.

[23] Heard, S.B. (1992). Patterns in tree balance among cladistic, phenetic, and randomly generated phylogenetic trees. *Evolution*, **46**, 1818-1826.

[24] Heard, S.B. (1996). Patterns in phylogenetic tree balance with variable and evolving speciation rates. *Evolution*, **50**, 2141-2148.

[25] Heard, S.B. and Hauser, D.L. (1995). Key evolutionary innovations and their ecological mechanisms. *Historical Biology*, **10**, 151-173.

[26] Heard, S.B. and Mooers, A.Ø. (2000). Phylogenetically patterned speciation rates and extinction risks change the loss of evolutionary history during extinctions. *Proceedings of the Royal Society of London Series B—Biological Sciences*, **267**, 613-620.

[27] Hein, J., Schierup, M. H., and Wiuf, C. (2005). *Gene Genealogies, Variation and Evolution*. Oxford University Press, Oxford.

[28] Hey, J. (1992). Using phylogenetic trees to study speciation and extinction. *Evolution*, **46**, 627-640.

[29] Hoelzer, G.A. and Melnick, D.J. (1994). Patterns of speciation and limits to phylogenetic resolution. *Trends in Ecology and Evolution*, **9**, 105-107.

[30] Holman, E. (2005). Nodes in phylogenetic trees: the relation between imbalance and number of descendent species. *Systematic Biology*, **54**, 895-899.

[31] Hubbell, S.P. (2001). *The Unified Neutral Theory of Biogeography*. Princeton University Press, Princeton.

[32] Hutchinson, G.E. (1959). Homage to Santa Rosalia, or why are there so many kinds of animals? *American Naturalist*, **93**, 145-159.

[33] Kingman, J.F.C. (1982). On the genealogy of large populations. *Journal of Applied Probability*, **19**, 27-43.

[34] Kinzig, A.P., Levin, S.A., Dushoff, J., and Pacala, S. (1999). Limiting similarity, species packing, and system stability for hierarchical competition-colonization models. *American Naturalist* , **153**, 371-383.

[35] Losos, J.B. and Adler, F.R. (1995). Stumped by trees? A generalized null

model for patterns of organismal diversity. *American Naturalist*, **145**, 329-342.

[36] Matsen, F. A. (2006). A geometric approach to tree shape statistics. *Systematic Biology*, **55**, 652-661.

[37] May, R.M. and MacArthur, R.H. (1972). Niche overlap as a function of environmental variability. *Proceedings of the National Academy of Science (USA)*, **69**, 1109-1113.

[38] Mayhew, P.J. (2002). Shifts in hexapod diversification and what Haldane could have said. *Proceedings of the Royal Society of London Series B—Biological Sciences*, **269**, 969-974.

[39] Mayr, E. (1954). Change of genetic environment and evolution. In *Evolution as a Process* (ed. J. Huxley, A. C. Hardy, and E. B. Ford), pp. 157-180. Allen and Unwin, London.

[40] McGill, B.J., Hadley, E.A., and Maurer, B.A. (2005). Community inertia of Quaternary small mammal assemblages in North America. *Proceedings of the National Academy of Science (USA)*, **102**, 16701-16706.

[41] Mooers, A.Ø. and Heard, S.B. (1997). Inferring evolutionary process from phylogenetic tree shape. *Quarterly Review of Biology*, **72**, 31-54.

[42] Moran, P.A.P. (1958). A general theory of the distribution of gene frequencies. 1. Overlapping generations. *Proceedings of the Royal Society of London Series B—Biological Sciences*, **149**, 102-112.

[43] Mitter, C., Farrell, B. and Wiegmann, B. (1988). The phylogenetic study of adaptive zones: Has phytophagy promoted insect diversification? *American Naturalist*, **132**, 107-128.

[44] Nee, S. (2001). Inferring speciation rates from phylogenies. *Evolution*, **55**, 661-668

[45] Nee, S. and May, R.M. (1997). Extinction and the loss of evolutionary history. *Science*, **278**, 692-694.

[46] Nee, S., Mooers, A.Ø., and Harvey, P.H. (1992). Tempo and mode of evolution revealed from molecular phylogenies. *Proceedings of the National Academy of Sciences (USA)*, **89**, 8322-8326.

[47] Nee, S., May, R.M., and Harvey, P.H. (1994). The reconstructed evolutionary process. *Philosophical Transactions of the Royal Society Series B—Biological Sciences*, **344**, 305-311.

[48] Nee, S., Holmes, E.C., May, R.M., and Harvey, P.H. (1995). Estimating extinction from molecular phylogenies. In *Estimating Extinction Rates* (ed. Lawton, J.L. and R.M. May), pp. 164-182. Oxford University Press.

[49] Nosil, P. and Mooers, A.Ø. (2005). Testing hypotheses about ecological specialization using phylogenetic trees. *Evolution*, **59**, 2256-2263.

[50] Paradis, E. (2005). Statistical analysis of diversification with species traits. *Evolution*, **59**, 1-12.

[51] Price T., Gibbs, H.L., de Sousa, L., and Richman, A.D. (1998). Different timing of the adaptive radiations of North American and Asian warblers. *Proceedings of the Royal Society of London Series B—Biological Sciences*,

265, 1969-1975.

[52] Purvis, A. (1996). Using interspecies phylogenies to test macroevolutionary hypotheses. In *New Uses for New Phylogenies.* (ed. P.H. Harvey, A.J.L. Brown, J.M. Smith and S. Nee), pp. 153-168. Oxford University Press, Oxford.

[53] Pybus, O.G and Harvey, P.H. (2000). Testing macro-evolutionary models using incomplete molecular phylogenies. *Proceedings of the Royal Society Series B—Biological Sciences*, **267**, 2267-2272.

[54] Pybus, O.G., Rambaut, A., Holmes, E.C., and Harvey, P.H. (2002). New inferences from tree shape: Numbers of missing taxa and population growth rates. *Systematic Biology*, **51**, 881-888.

[55] Rabosky, D.L. (2006). Likelihood methods for detecting temporal shifts in diversification rates. *Evolution*, **60**, 1152-1164.

[56] Raup, D.M., Gould, S.J., Schopf, T.J.M., and Simberloff, D.S. (1973). Stochastic models of phylogeny and the evolution of diversity. *Journal of Geology*, **81**, 525-542.

[57] Rauch, E.M. and Bar-Yam, Y. (2004). Theory predicts the uneven distribution of genetic diversity within species. *Nature*, **431**, 449-452.

[58] Rogers, J.S. (1994). Central moments and probability-distribution of Colless' Coefficent of tree imbalance. *Evolution*, **48**, 2026-2036.

[59] Rohlf, F.J., Chang, W.S., Sokal, R.R., and Kim, J. (1990). Accuracy of estimated phylogenies: effects of tree topology and evolutionary model. *Evolution*, **44**, 1671-1684.

[60] Rüber, L., and Zardoya, R. (2005). Rapid cladogenesis in marine fishes revisited. *Evolution*, **59**, 1119-1127.

[61] Savolainen, V., Heard, S.B., Powell, M.P., Davies, T.J., and Mooers A.Ø. (2002). Is cladogenesis heritable? *Systematic Biology*, **51**, 835-843.

[62] Schluter, D. (2000). *Ecology of Adaptive Radiation.* Oxford University Press, Oxford, UK.

[63] Schneider H., Schuettpelz E., Pryer K.M., Cranfill R., Magallon S., and Lupia R. (2004) Ferns diversified in the shadow of angiosperms *Nature* **428**, 553-557.

[64] Sepkoski, J.J., Jr. (1979). A kinetic model of Phanerozoic taxonomic diversity. II. Early Phanerozoic families and multiple equilibria . *Paleobiology*, **5**, 22-251.

[65] Sepkoski, J.J., Jr. (1984). A kinetic model of Phanerozoic taxonomic diversity. III. Post-Paleozoic families and mass extinction. *Paleobiology*, **10**, 246-267.

[66] Shaw, A.J., Cox, C.J., Goffinet, B., Buck, W.R., and Boles, S.B. (2003). Phylogenetic evidence of a rapid radiation of pleurocarpous mosses (Bryophyta). *Evolution*, **57**, 2226-2241.

[67] Sims, H.J. and McConway, K.J. (2003). Nonstochastic variation of species-level diversification rates within angiosperms. *Evolution*, **57**, 460-479.

[68] Simberloff, D.S., Hecht, K.L., McCoy, E.D., and Conner, E.F. (1981). There

have been no statistical tests of cladistic biogeographic hypotheses. In *Vicariance Biogeography: a critique* (ed. G. Nelson and D. E. Rosen), pp. 40-63. Columbia University Press, New York.

[69] Stam, E. (2002). Does imbalance in phylogenies reflect only bias? *Evolution*, **56**, 1292-1295.

[70] Steel, M. and McKenzie, A. (2001). Properties of phylogenetic trees generated by Yule-type speciation models. *Mathematical Biosciences*, **170**, 91-112.

[71] Stireman, J.O. III., Nason, J.D., Heard, S.B., and Seehawer, J.M. (2006). Cascading host-associated genetic differentiation in parasitoids of phytophagous insects. *Proceedings of the Royal Society of London Series B—Biological Sciences*, **273**, 523-530.

[72] Strathman, R.R. and Slatkin, M. (1983). The improbability of animal phyla with few species. *Paleobiology*, **9**, 97-106.

[73] Volkov, I., Banavar, J.R., Hubbell, S.P., and Maritan, A. (2003). Neutral theory and relative species abundance in ecology. *Nature*, **424**, 1035-1037.

[74] Vrba, E.S. and Eldredge, N. (1984). Individuals, hierarchies, and processes: towards a more complete evolutionary theory. *Paleobiology*, **10**, 146-171.

[75] Wilkinson, M., Cotton, J.A., Creevey, C., Eulenstein, O., Harris, S.R., Lapointe, F.J., Levasseur, C., McInerney, J.O., Pisani, D., and Thorley, J.L. (2005). The shape of supertrees to come: Tree shape related properties of fourteen supertree methods. *Systematic Biology*, **54**, 419-431.

[76] Yang, Z. and Rannala, B. (1997). Bayesian phylogenetic inference using DNA sequences: a Markov Chain Monte Carlo Method. *Molecular Biology and Evolution*, **14**,717-724.

[77] Yule, G.U. (1924). A mathematical theory of evolution based on the conclusions of Dr J. C. Willis. *Philosophical Transactions of the Royal Society (London) Series B—Biological Sciences*, **213**, 21-87.

[78] Zhaxybayeva, O. and Gogarten, J.P. (2004). Cladogenesis, coalescence and the evolution of the three domains of life. *Trends in Genetics*, **20**, 182-187.

[79] Zink, R.M. and Slowinski, J.B. (1995). Evidence from Molecular Systematics for Decreased Avian Diversification in the Pleistocene Epoch. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, **92**, 5832-5835.