THE ROYAL
SOCIETY

Received 7 January 2003
Accepted 18 February 2003
Published online 20 May 2003

Evolutionary models of phylogenetic trees

Iosif Pinelis

Department of Mathematical Sciences, Michigan Technological University, Houghton, MI 49931, USA (*ipinelis@mtu.edu*)

The most widely used evolutionary model for phylogenetic trees is the equal-rates Markov (ERM) model. A problem is that the ERM model predicts less imbalance than observed for trees inferred from real data; in fact, the observed imbalance tends to fall between the values predicted by the ERM model and those predicted by the proportional-to-distinguishable-arrangements (PDA) model. Here, a continuous multi-rate (MR) family of evolutionary models is presented which contains entire subfamilies corresponding to both the PDA and ERM models. Furthermore, this MR family covers an entire range from 'completely balanced' to 'completely unbalanced' models. In particular, the MR family contains other known evolutionary models. The MR family is very versatile and virtually free of assumptions on the character of evolution; yet it is highly susceptible to rigorous analyses. In particular, such analyses help to uncover adaptability, quasi-stabilization and prolonged stasis as major possible causes of the imbalance. However, the MR model is functionally simple and requires only three parameters to reproduce the observed imbalance.

Keywords: evolutionary models; imbalance of phylogenetic trees; proportional-to-distinguishable-arrangements model; adaptability; quasi-stabilization; prolonged stasis

1. INTRODUCTION

The most widely used evolutionary model for phylogenetic trees is the so-called equal-rates Markov (ERM) model, which is based on a pure-birth branching process (Yule 1924; Kendall 1948). However, there is a problem in that the ERM model predicts less imbalance between the sizes of the daughter clades of a species than observed for trees inferred from real data. In fact, the observed imbalance of phylogenetic trees tends to fall between the values predicted by the ERM model and those predicted by the so-called proportional-to-distinguishable-arrangements (PDA) model (see Aldous (1996, 2001) and http://stat-www.berkeley.edu/users/aldous/csuros.html; Guyer & Slowinski 1991; Heard 1992; Rogers 1994, fig. 5; Mooers 1995; Mooers & Heard 1997, fig. 4).

However, an evolutionary interpretation for the PDA model has been given only recently by Steel & McKenzie (2001; cf. Simberloff *et al.* 1981; Slowinski 1990; Guyer & Slowinski 1991, 1993; Maddison & Slatkin 1991; Heard 1992; Nee *et al.* 1992; Rogers 1993; Cunningham 1995; Aldous 1996, 2001; Mooers & Heard 1997; Harcourt-Brown *et al.* 2001).

This paper provides a general evolutionary interpretation of the PDA model. (An evolutionary model is understood here as one which describes a gradual development of phylogenetic trees in time, from the root species onwards.)

Moreover, a continuous multi-rate (MR) family of evolutionary models is presented which contains both the PDA and ERM models. In particular, it will follow that the recent conjecture by Harcourt-Brown *et al.* (2001)—that the so-called ERM-TI ('time-inclusive') model and the PDA model are mathematically equivalent to each other—is almost (but not quite) true.

Furthermore, the MR family covers an entire range from 'completely balanced' (CB) to 'completely unbalanced' (CU) models. In particular, the MR family contains the evolutionary models proposed by Losos & Adler (1995), Heard (1996) and Steel & McKenzie (2001).

The MR family is very versatile and virtually free of assumptions on the character of evolution, yet it is highly susceptible to rigorous analyses. In particular, such analyses help to uncover adaptability, quasi-stabilization and prolonged stasis as major possible causes of the imbalance. However, the MR model is functionally simple and requires only three parameters to reproduce the observed imbalance.

2. A CONVENIENT (IF UNUSUAL) NOTION OF THE PHYLOGENETIC TREE

If the particular names of the species are considered irrelevant and are each replaced by one symbol, say $*$, then a (finite non-empty rooted dichotomous phylogenetic) *tree* and its *size* can be defined recursively as follows.

(i) Every tree $\mathfrak{t}$ has a definite size, denoted here by $|\mathfrak{t}|$, which is a natural number.
(ii) The only tree of size 1 is $*$.
(iii) For any natural number $w \geqslant 2$, $\mathfrak{t}$ is a tree of size w if and only if $\mathfrak{t}$ is the ordered pair $(\mathfrak{p}, \mathfrak{d})$ of some trees $\mathfrak{p}$ and $\mathfrak{d}$ of smaller sizes, satisfying the condition $|\mathfrak{p}| + |\mathfrak{d}| = w$.

By definition, two ordered pairs $(\mathfrak{p}, \mathfrak{d})$ and $(\tilde{\mathfrak{p}}, \tilde{\mathfrak{d}})$ are considered the same if and only if $\mathfrak{p} = \tilde{\mathfrak{p}}$ and $\mathfrak{d} = \tilde{\mathfrak{d}}$.

Thus, here the trees $(\mathfrak{p}, \mathfrak{d})$ and $(\mathfrak{d}, \mathfrak{p})$ are considered to be different if $\mathfrak{p} \neq \mathfrak{d}$. If $\mathfrak{t} = (\mathfrak{p}, \mathfrak{d})$, then $\mathfrak{p}$ and $\mathfrak{d}$ will be referred to, respectively, as the parent and daughter branches of tree $\mathfrak{t}$.

It is sometimes convenient if the notion of a finite *non-empty tree* is extended to include also the empty tree, of

DOI 10.1098/rspb.2003.2374

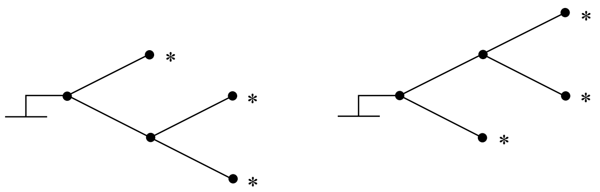

Figure 1. Trees of size 3.

size 0, denoted here by $\varnothing$, and the infinite trees, of size ∞.

The (topological) *type* $\mathscr{T}$ (also referred to as the *shape*) of any finite tree t of size $\geqslant 2$ can be obtained by replacing all the ordered pairs $(\mathfrak{p}, \mathfrak{d})$ in the representation of t by the corresponding non-ordered pairs $[\![\mathscr{P}, \mathscr{D}]\!]$. Two non-ordered pairs $[\![\mathscr{P}, \mathscr{D}]\!]$ and $[\![\tilde{\mathscr{P}}, \tilde{\mathscr{D}}]\!]$ are considered the same if and only if either $\mathscr{P} = \tilde{\mathscr{P}}$ and $\mathscr{D} = \tilde{\mathscr{D}}$, or $\mathscr{P} = \tilde{\mathscr{D}}$ and $\mathscr{D} = \tilde{\mathscr{P}}$.

For example, $(*, (*, *))$ and $((*, *), *)$ are the only trees of size 3. For the first of them, the parent branch is $*$ and the daughter branch is $(*, *)$; for the second tree, it is vice versa. Both of these trees are of type $[\![*, [\![*, *]\!]]\!]$, which is the same as $[\![[\![*, *]\!], *]\!]$. Figure 1 illustrates the graphs of these two trees.

Also, there are five trees of size 4: $(*, (*, (*, *)))$, $(*, ((*, *), *))$, $((*, (*, *)), *)$, $(((*, *), *), *)$ and $((*, *),(*, *))$; the first four of them are of tree type $[\![*, [\![*, [\![*, *]\!]]\!]]\!]$ and the fifth, of type $[\![[\![*, *]\!], [\![*, *]\!]]\!]$.

Let us denote the types of the trees $\varnothing$ and $*$ by the same symbols, $\varnothing$ and $*$, respectively.

In the above definition of the tree, not only the names of the species, but also their ages (i.e. the lengths of the edges) are ignored. However, we shall see that in our general MR model, the age may well be one of the attributes of the state of the species under consideration.

Apparently, the crucial ingredient of our approach is the observation that the usual definition of the PDA model (Rosen 1978; Mooers & Heard 1997) is equivalent (in terms of the conditional distribution of the tree type given the size) to the condition that all trees of any given size be *equally probable*; for details, see § 1 of electronic Appendix A (available at The Royal Society's Publications Web site). This observation, while simple, does not seem to be completely trivial; indeed, it provides the essential and rather subtle link. The reason is that the above definition of a *tree* (which is more convenient for establishing equation (3.1)) is different from the definition of a *distinguishable arrangement* (DA) normally used to define the PDA model. Moreover, as explained in § 1 of electronic Appendix A, in general there can be no correspondence between the trees and the DAs of the same size such that the same number of DAs correspond to every tree (or vice versa). There are, for example, *two trees* of size 3 versus *three* DAs of any given set of three species. The very notion of a tree, defined above as if proceeding from the root onwards, appears more amenable to an evolutionary interpretation than the notion of a DA, which 'proceeds' from a given set of species back in time to the root; by contrast, in real evolutionary processes the resulting set of species is not given in advance.

3. GENERAL EVOLUTIONARY CONDITIONS FOR THE PDA DISTRIBUTION

Let $\mathfrak{T}$ denote the random tree that 'grows' in the evolutionary model under consideration; $\mathfrak{T}$ may, with a zero or non-zero probability, be empty or infinite. Note that the term 'random' is used in this paper in the general sense assumed in probability theory; in particular, this term does not necessarily imply that the corresponding probability distribution is uniform in some sense. In most contexts in this paper, the term 'random tree' will actually mean 'the tree effected by a random evolutionary mechanism'.

Let us assume that the evolution begins with a single species, which is at the root of the random tree $\mathfrak{T}$.

Assume that any species may give birth, if ever, only to one species at a time.

Let FNE stand for the event that the random tree $\mathfrak{T}$ happens to be *finite* and *non-empty*.

Let FB stand for the event (*of finite branching*) that $\mathfrak{T}$ happens to be finite and of size $\geqslant 2$, so that $\mathfrak{T}$ is the ordered pair $(\mathfrak{P}, \mathfrak{D})$ of finite non-empty trees $\mathfrak{P}$ and $\mathfrak{D}$ (the two branches of $\mathfrak{T}$, the parent and daughter ones).

For any given non-random finite non-empty tree t, let $\hat{p}(\mathsf{t})$ stand for the conditional probability that the random tree $\mathfrak{T}$ happens to coincide with t given FNE; let $p(\mathsf{t})$ stand for the corresponding unconditional probability.

Let $\hat{s}$ denote the conditional probability $\mathrm{P}(\mathrm{FB}\,|\,\mathrm{FNE})$ of FB given FNE.

Introduce the following conditions:

(i) (PDA1): the probability P(FB) is non-zero;

(ii) (PDA2): given FB, the parent and daughter branches $\mathfrak{P}$ and $\mathfrak{D}$ of the random tree $\mathfrak{T}$ are conditionally independent of each other; moreover, the conditional distribution of each branch given FB is the same as the conditional distribution of the entire tree $\mathfrak{T}$ given that $\mathfrak{T}$ is finite and non-empty.

Thus, condition (PDA2) implies that the parent and daughter branches $\mathfrak{P}$ and $\mathfrak{D}$ have the same conditional probability distribution (which, of course, does not mean that the two branches must always be the same or even of the same size).

Conditions (PDA1) and (PDA2) imply $\hat{p}(\mathsf{t}) = \hat{s}\hat{p}(\mathfrak{p})\hat{p}(\mathfrak{d})$ for any (non-random finite) tree t of the form $(\mathfrak{p}, \mathfrak{d})$; this can be rewritten in terms of the unconditional probabilities as follows:

$$p(\mathsf{t}) = sp(\mathfrak{p})\,p(\mathfrak{d}) \text{ if } \mathsf{t} = (\mathfrak{p}, \mathfrak{d}), \tag{3.1}$$

where $s := \hat{s}/\mathrm{P}(\mathrm{FNE})$. In general, s cannot be interpreted as a probability; indeed, s may be greater than 1 (see the note immediately after equation (16) in electronic Appendix A). Now it follows by induction that

$$p(\mathsf{t}) = s^{|\mathsf{t}|-1}p(*)^{|\mathsf{t}|} \tag{3.2}$$

for any finite non-empty tree t. Hence, all the trees t of any given finite size $|\mathsf{t}|$ are equally probable, and their probabilities are non-zero. As mentioned earlier, this equiprobability condition determines the PDA model.

Thus, very simply, one obtains a general evolutionary interpretation of the PDA model—provided conditions (PDA1) and (PDA2) apply. It seems that this simplicity is mainly due to the established equivalence (in terms of the distribution of the tree type) of the notion of the DA to that of the tree.

Both conditions (PDA1) and (PDA2) may seem rather natural, so much so that now one may begin to wonder, not why the observed imbalance is biased from the ERM values up towards the PDA ones, but why the observed imbalance values are usually significantly below those predicted by the PDA.

Indeed, condition (PDA1) may seem rather trivial. As for condition (PDA2), according to Guyer & Slowinski (1993), it appears to be approximately satisfied, at least for large (and apparently old) trees.

However, at this point one can see a problem if condition (PDA2) is to be satisfied *exactly*: in order for the parent and daughter branches of the random tree $\mathfrak{T}$ to be *exact* copies of $\mathfrak{T}$ *in distribution*, $\mathfrak{T}$ seems to have to be infinitely old, at least if the rates are constant in time. However, then there may be a problem with condition (PDA1). For example, if the random tree $\mathfrak{T}$ is modelled by the live particles in the simple birth-and-death process (BDP) over the entire time interval from 0 to ∞, then the probability $p(*)$ is always zero—as well as the probability $p(t)$ of any finite non-empty tree t; indeed, with a probability of 1, the number of live particles in the BDP tends either to 0 or to ∞ as the age of the tree goes to ∞.

For model trees $\mathfrak{T}$ that are actually infinitely old, there are then at least two ways to make the probability $p(*)$ non-zero: either include the extinct species into the tree and/or assume that the species may stabilize in time, so that their rate of change eventually comes down to zero. Of course, the latter condition can hardly ever be assumed to be satisfied exactly, and that is one reason why exact PDA values of imbalance are hardly ever observed if only the extant species and their lineages are included into the tree.

However, it is clear from the above that the greater the extent to which conditions (PDA1) and (PDA2) are satisfied, the closer will be the observed imbalance $\mathfrak{T}$ to the PDA value.

Thus, one can now explain the fact that the observed imbalance is biased upwards from the ERM model values towards the PDA values—by suggesting that the reason for such bias is that conditions (PDA1) and (PDA2) are often satisfied, at least to some extent.

In § 3 of electronic Appendix A, it will be made clear that conditions (PDA1) and (PDA2) are indeed satisfied in the MR model when certain additional conditions on the parameters of the MR model hold. One may say that conditions (PDA1) and (PDA2) are not quite evolutionary, in that they are, to a certain degree, anticipatory of the size of the entire random tree $\mathfrak{T}$. However, neither the subsequent formulation of the MR model nor the conditions on its parameters that will imply (PDA1) and (PDA2) will be anticipatory to any extent; rather, they will be stated in purely evolutionary terms.

One simple (if not quite realistic) case in which conditions (PDA1) and (PDA2) are obviously satisfied is when the following conditions hold: (i) extinction is not allowed; (ii) the speciation rate depends only on the age of a species; (iii) any species may split into two *only before* it reaches a certain age (say A); and (iv) the age of the tree $\mathfrak{T}$ is greater than $|\mathfrak{T}|\ A$ (so that the tree cannot grow any more). In this case, Steel & McKenzie (2001) showed (in a different manner) that one has an instant of the PDA model. We shall show (in § 5 of electronic Appendix A) that this case is also a special case of the MR model.

4. MULTI-RATE EVOLUTIONARY MODEL: DESCRIPTION, DISCUSSION AND SIMPLEST SPECIAL CASES

Simulation studies (Heard 1996) demonstrate that large enough diversification rate changes at speciation events may account for observed amounts of imbalance.

In general accordance with such studies, let us define here what may be called the MR model.

For a random (phylogenetic) tree, let S denote the state space—that is the set of *states* of actually or potentially existing 'species'; the latter term may stand for species proper, as well as (possibly geographically separated) subspecies or other taxa. Any state i in S may carry any kinds of information about the (sub)species, such as its identifier, genotype, age, size, geographical location, feeding and behavioural patterns, the state of competition and available resources.

To avoid non-essential technical complications, let us assume that the state space S is countable (by some or all natural numbers); thus, S is allowed to be infinite.

Next, let us assume that the evolution of the tree begins at time $t = 0$ with a single species at the root of the tree. It proceeds as follows: in any infinitely small time interval $(t, t + \mathrm{d}t)$, any species which was in a state i at time $t \geqslant 0$,

(i) with probability $\mu_{ij}\mathrm{d}t$, becomes transformed into some other state j

(ii) or, with probability $\sigma_{ij}\mathrm{d}t$, remains unchanged and at that gives birth to a separate species which is born in some state j

(iii) or, with probability $1 - \lambda_i\mathrm{d}t$, does not take part in any phylogenetic change (i.e. its state remains unchanged and it does not give birth to any separate species); here and in what follows

$$\lambda_i := \mu_{iS} + \sigma_{iS} \tag{4.1}$$

and

$$\mu_{iA} := \sum_{j \in A} \mu_{ij} \text{ and } \sigma_{iA} := \sum_{j \in A} \sigma_{ij},$$

for any subset A of the state space S.

It is naturally assumed that, for any states i and j, μ_{ij} and σ_{ij} are non-negative numbers; they may be referred to, respectively, as the transformation and speciation rates. These rates may also depend on time t. However, without loss of generality, the rates will be assumed to be 'constant', because the current time moment itself can be considered as one of the attributes of the state of the species (cf. the consideration of the LA-AD and LA-RAD models in § 5 of electronic Appendix A). Also, it is assumed that for all i in S

$$\mu_{ii} = 0;$$

that is, any 'transformation' of a state into itself is not counted as a bona fide transformation.

All of the species at any given time moment are evolving independently of one another and of the prehistory.

For each state i, the number λ_i may be interpreted as the total rate of phylogenetic change of a species in state i, so that $\lambda_i \mathrm{d}t$ is the probability that a species which was in state i at a time moment t takes part in any phylogenetic change in the time interval $(t, t + \mathrm{d}t)$.

Mathematically, the MR model is a modification of a multi-type branching process (cf. Harris 1963) with at most two offspring of one particle at a time; the modification here consists in making a distinction in some instances between the parent and daughter species.

Transformations of a (sub)species from one state into another may include retainable mutations as well as any other changes in the attributes of the state of the (sub)species: the geographical location, age, size, feeding and behavioural patterns, etc. Geographical separation of subspecies may also cause a split, that is, the birth of a separate species.

In addition to the choice of the parameters μ_{ij} and σ_{ij}, there are three other important choices that one can make when working with an MR model.

First, one can choose the *stopping time* $T > 0$, so that $[0, T]$ is the time interval over which the branching process is observed. However, it is not difficult to see that the distribution of the random tree $\mathfrak{T}$ in an MR model depends on the μ_{ij}s, σ_{ij}s and T only through the products $\mu_{ij} T$ and $\sigma_{ij} T$. Thus, the significant choice of T is mainly between the cases $T < \infty$ and $T = \infty$. It will be shown that the PDA model results from the MR model in the latter of these two cases. The stopping time T may also be random; for example, for any natural number w, one may choose $T = T_w$, where T_w stands for the first time moment when the size of the random tree becomes equal to w.

Second, for any subset A of S, one may choose to exclude from the random tree $\mathfrak{T}$ the species followed in time *only* by species whose state at time T belongs to the set A. In such a case, let us refer to the random tree $\mathfrak{T}$ and the corresponding MR model as *A-incomplete*; let us use the terms *incomplete* and *complete* to mean, respectively, 'A-incomplete for some non-empty set A' and 'not incomplete'. For example, if D stands for the set of the states of the extinct species, then a D-incomplete random tree $\mathfrak{T}$ will contain only the lineages of the species extant at time T.

Third, one can choose an initial state distribution on S, i.e. the probability distribution of the state of the root species at time $t = 0$.

The ERM model (without extinction) may be considered as the simplest example of a complete MR model, with the state space S consisting just of one state (say $i = 1$), $\mu_{11} = 0$, $\sigma_{11} = \lambda > 0$, and $T = T_w$ or a non-random $T < \infty$ (see § 3.1 in electronic Appendix A concerning a relation between these two choices of T).

The birth-and-death (branching) process (BDP) (see Kendall 1948) with rates λ and μ of birth and death (respectively) is mathematically equivalent to another simple example of an MR model, with the state space S consisting of two states, say labelled by 0 and 1, with $\mu_{10} = \mu > 0$, $\mu_{11} = 0$, $\sigma_{10} = 0$, $\sigma_{11} = \lambda > 0$, and $\mu_{00} = \mu_{01} = \sigma_{00} = \sigma_{01} = 0$. This model is complete or $\{0\}$-incomplete depending on whether the dead particles and their lineages are included or not included into the random tree $\mathfrak{T}$. The dead particles (i.e. those in state 0) here may model either the extinct species or the quasi-stable ones, which are introduced and discussed in § 5.

The phenomenon of extinction was considered, for example, in Gould *et al.* (1977), Stanley *et al.* (1981), Nee *et al.* (1992, 1994*a*,*b*, 1995, 2001), Mooers & Heard (1997) and Heard & Mooers (2002).

The ERM-TI model studied by Harcourt-Brown *et al.* (2001) corresponds to the special case of the complete BDP random tree with $\lambda = \mu$ and $T = T_w$.

5. MR INTERPRETATIONS OF THE (S)PDA AND ERM MODELS

It will be verified in § 3 of electronic Appendix A that the PDA model can be interpreted as a special case (or, rather, as an entire variety of cases) of the MR model. Any such interpretation will be called here an MR-PDA interpretation (or model). Similar MR interpretations will be given to the ERM model and to what is referred here to as the super-PDA (SPDA) models, even more 'unbalanced' than the PDA model; an extreme case of an SPDA model is a CU model, in which the random evolutionary tree $\mathfrak{T}$ is CU with probability 1; that is, with probability 1, the type of $\mathfrak{T}$ is one of the CU types: $\varnothing, *, [\![*, *]\!], [\![*, [\![*, *]\!]]\!], [\![*, [\![*, [\![*, *]\!]]\!]]\!], \ldots$. Let us refer to such MR interpretations, respectively, as MR-ERM, MR-SPDA and MR-CU interpretations (or models).

Suppose that the state space S can be partitioned into three subsets of it, denoted here by D, U and Q; in principle, some of these subsets may be empty.

The set D consists of the states of the *dead* (i.e. extinct) species. This implies that for the total rates of change λ_i given by equation (4.1) one has

$\lambda_i = 0$ for every state i in D.

To simplify the account, let us also make the rather natural assumption that

$$\sigma_{iD} = 0 \text{ for every state } i \text{ in } S; \qquad (5.1)$$

that is, any birth of a stillborn species (extinct immediately after its birth) is not counted as a bona fide birth.

The set U consists of the (essentially) *unstable* states of species. Let us then make the natural assumption that

$\sigma_{iS} > 0$ at least for one state i in U;

that is, in at least one unstable state, a species may give birth to another species with a non-zero probability. Just to simplify the presentation, let us also assume that the set U of unstable states is *finite* and all of the unstable states *communicate* with one another; that is, any unstable state can be reached from any other unstable state in a finite time with a non-zero probability; to that end, it would be more than enough to assume that $\mu_{ij} > 0$ for all i and j in U such that $j \neq i$. It can then be demonstrated that there exists a unique probability distribution $(\pi_i)_{i \in U}$ on the set U with the following important property:

> take any time moment $t > 0$; then, given that the probability distribution of the state of a species at time moment zero is $(\pi_i)_{i \in U}$ and before time t the species does not give birth to a separate species and its state remains in the set U—the conditional distribution of the state of the species at time t will still be $(\pi_i)_{i \in U}$.

See details in § 2 of electronic Appendix A.

Let us refer to the probability distribution $(\pi_i)_{i \in U}$ as the *conditionally stationary distribution* (on the set U of all unstable species). In the trivial case when the set U is a singleton, there is only one probability distribution on U, and it is then necessarily and obviously the conditionally stationary distribution.

The set Q consists of the *quasi-stable* states of the *extant* species. It is assumed that

$\sigma_{iS} = \mu_{iU} = \mu_{iD} = 0$ for every state i in Q;

this means that any species in any quasi-stable state can only switch to another quasi-stable state; it cannot give birth or switch to an unstable state or become extinct.

It will be seen in §§ 3.2 and 3.3 of electronic Appendix A that the notion of quasi-stable states is crucial for our evolutionary interpretations of the PDA and SPDA models (cf. the discussion in § 3, above).

The set Q of quasi-stable states may be infinite; it also may be empty. Species in quasi-stable states may be thought of as extremely adaptable ones. Typically, they are eternal nomads, wandering (not just geographically but in other 'dimensions' as well) within the confines of the subset Q of the state space S. They may be changing attributes of their state: the geographical location, age, size, feeding and behavioural patterns, even the genotype, to a certain extent. However, they never become extinct or give birth to another species.

Of course, the notion of quasi-stability is an abstraction; in reality, instead of being such bona fide eternal nomads, some species may conform with this quasi-stability pattern only for a more or less prolonged period of time.

A very particular instance of a quasi-stable state is a (bona fide) *stable* state, which never changes. However, let us emphasize that stable states are *not* needed for the purposes of modelling presented in this paper.

However (while bona fide stabilization seems to be considered quite atypical among specialists on the evolutionary tree shape), the notion of *prolonged stasis* (possibly intermittent with periods of rapid speciation) is assumed by many other specialists as given, and the only problem for them there is how to explain such a phenomenon. For example, a query 'KW = (evolution∗ AND stasis∗)' in the database Biological Sciences of Cambridge Scientific Abstracts returns 134 records. For instance, articles by Soltis *et al.* (2002), Tamas *et al.* (2002) and Wernegreen (2002) (which are records 1, 4 and 6 of the 134) discuss 'genome stasis over the 50–70 million years of their evolution' and '"molecular living fossils", consistent with their relative morphological stasis for the past 165–200 million years'.

In this respect, note that, in particular, models such as those near the left endpoint $\alpha = 0$ of the α-spectrum of MR models presented in § 4 of electronic Appendix A may adequately represent the phenomenon of periods of prolonged stasis alternating with short time-periods of rapid speciation.

This resembles the Second Law of Thermodynamics, where the 'stabilization' phenomenon of 'steadily' and 'irreversibly' increasing entropy can be adequately explained by time-reversible processes; there too, the periods of instability and 'coming back to life' are relatively short, yet they forever alternate with prolonged 'stabilization' periods.

Note, also, that in the model referred to of Steel & McKenzie (2001), bona fide stabilization is *certain to occur* for *every* species upon its reaching a certain age A. By contrast, in the general MR model only quasi- (rather than bona fide) stabilization is needed for an evolutionary interpretation of the PDA model, and we do not require that even quasi-stabilization be certain to occur for any given species. Another significant difference is that our MR interpretations of the PDA model incorporate the entinction phenomenon as well. (Taking extinction into account does not change, in a certain sense, the tree-shape distribution—under the ERM-type condition that both the speciation and extinction rates are the same for all species at any given time moment; see Slowinski & Guyer (1989) and Rogers (1994). However, extinction does affect the tree-shape distribution if the equal-rates condition fails; see Heard & Mooers 2002.)

In the remainder of this section, the following conditions are assumed, in addition to the ones stated above.

(i) (MR1): the random tree $\mathfrak{T}$ is either complete or D-incomplete;

(ii) (MR2): the probability distribution of the state of the root species at time $t = 0$ coincides with the conditionally stationary distribution $(\pi_i)_{i \in U}$ on U;

(iii) (MR3): $\sigma_{ij} = \sigma_{iU}\pi_j$ for all i and j in U.

Condition (MR1) means that the extinct species and their lineages may or may not be included into the random tree $\mathfrak{T}$, while all species in all the other states (as well as their lineages) are included.

Condition (MR3) means that the conditional state distribution of the daughter species upon its birth coincides with the conditionally stationary distribution $(\pi_i)_{i \in U}$ on U—given that the birth event with the parent species in state i occurs *and* the daughter species is in an unstable state upon its birth.

As will be explained in § 2 of electronic Appendix A (in the paragraph following equation (4)), conditions (MR2) and (MR3) make sense, because, under mild general restrictions, the stationary distribution is the limit one.

In § 3 of electronic Appendix A it will be shown that, under the general conditions stated above in this section, MR-ERM, MR-PDA and MR-CU models result depending only on whether σ_{iU}, μ_{iQ} and σ_{iQ} are zero or not and whether T is finite or not, according to table 1; thus, vastly different models result from rather subtle differences in the conditions.

In table 1, '= 0' means '= 0 for all i in U' and '> 0' means '> 0 at least for one i in U'; '$\geqslant 0$' means that it does not matter whether or not for some or all i in U the (nonnegative) number is zero or not; $T < \infty$ means that either T is a non-random finite positive number or $T = T_w$ for some natural number w; $T \leqslant \infty$ means that any of the three possibilities are valid: $T = \infty$ or T is a non-random finite positive number or $T = T_w$ for some natural number w.

Speaking somewhat loosely, according to table 1 and under conditions (MR1)–(MR3), an MR-ERM model results if the stopping time T is finite and quasi-stabilization is certain never to occur, whether by transformation

Table 1. Sufficient conditions for MR-ERM, MR-PDA and MR-CU models.

parameters				
σ_{iU}	μ_{iQ}	σ_{iQ}	T	resulting model
> 0	$= 0$	$= 0$	$< \infty$	MR-ERM
> 0	> 0	$= 0$	$= \infty$	MR-PDA
$= 0$	$\geqslant 0$	> 0	$\leqslant \infty$	MR-CU

or by birth. Next, an MR-PDA model results if T is infinite and quasi-stabilization does occur with a non-zero probability—but only by transformation and not by birth. (In such MR-ERM and MR-PDA models, speciation occurs with a non-zero probability.) Finally, an MR-CU model results if only quasi-stable species may be born with a non-zero probability.

Thus, the MR family contains an entire subfamily of evolutionary interpretations for each of these models: ERM, PDA and CU; such subfamilies arise for every given number $\geqslant 2$ of states in the state space S.

Being continuous, the MR family also contains continuous *one-parameter* spectra of evolutionary models interpolating from an ERM model to a PDA one to a CU one. Along such spectra, one proceeds, as it were, from the relatively small imbalance in the ERM models to the larger imbalance in the PDA models to the largest possible imbalance in the CU models. Simple examples of such spectra are given in §§ 3.3 and 4 of electronic Appendix A.

Other continuous spectra of models interpolating between the ERM and PDA models were given (Aldous 1996, 2001). Also, Cunningham (1995) discussed, briefly and informally, a range of null models between the ERM and PDA ones. However, those models of Aldous and Cunningham did not have an evolutionary interpretation.

The simplest example of an MR-PDA model is the 'BDP' case of the MR model, mentioned at the end of § 4, where $T = \infty$, $S = \{0, 1\}$, $U = \{1\}$, $Q = \{0\}$ and $D = \varnothing$, so that state 0 is interpreted here as a (quasi-)stable one.

Alternatively, as indicated in § 4, one can set here $U = \{1\}$, $Q = \varnothing$, and $D = \{0\}$ and require that the random tree $\mathfrak{T}$ be complete; thus, state 0 will be interpreted now as that of being extinct; nonetheless, the extinct species and their lineages will be included into the tree.

The special case (with $\lambda = \mu$) of the latter alternative corresponds to the ERM-TI model studied by means of computer simulation by Harcourt-Brown *et al.* (2001)—except that they had $T = T_w$ instead of $T = \infty$. Based on the computer simulation, they proposed that the ERM-TI model is mathematically equivalent to the PDA model, even though these two models were 'seemingly very different' to them. Now it follows that their conjecture is almost true—only one has to use $T = \infty$ instead of $T = T_w$.

However, if the extant species are interpreted as only the live particles in the BDP, then, according to table 1, one will have an MR-ERM model provided also that $T < \infty$; for $T = T_w$, this result was obtained by Slowinski & Guyer (1989) and Rogers (1994).

Let us emphasize again that in this paper an evolutionary interpretation is understood merely as one which describes a gradual development of phylogenetic trees in time, from the root onwards. Thus, the statement that the PDA or CU model is given an evolutionary interpretation does not imply that the PDA or CU model perfectly describes all observed phylogenetic trees. Actually, that is rarely (if ever) the case for the PDA model (especially if only the lineages of the extant species are considered), and practically never the case for the CU model. Even in general, it should be clear that there hardly can be a single model which would *perfectly* describe all evolutionary phenomena.

However, MR models intermediate between MR-ERM and MR-PDA may well be rather adequate. See, for example, the continuous α-spectra of MR models described in § 4 of electronic Appendix A. For any such α-spectrum, the set Q of quasi-stable states is empty for all values of the interpolation parameter α in the interval $[0, 1]$ except for the limit value $\alpha = 0$.

Note that the above conditions for MR-ERM, MR-PDA and MR-CU models are merely sufficient, not necessary. For example, an alternative way to obtain D-incomplete MR-ERM models is to replace conditions (MR2) and (MR3) by the conditions that the σ_{iU}s and the μ_{iD}s do not depend on i in U and the state at time 0 be unstable (at that, the μ_{iU}s may well depend on i in U); indeed, in this case one can merge all the states in U into one state without altering the distribution of the random tree.

6. OTHER WORK ON EVOLUTIONARY MODELS WITH VARYING RATES

Losos & Adler (1995) studied a modification of the ERM model in which 'the length [...] of the speciation process, the refractory period', is assumed to be non-zero. The effect of this modification is that the imbalance value in the LA model is generally not between those in the ERM and PDA model. Rather, in contrast with observed values, the imbalance values in the LA model are less than those in the ERM model (and hence less than those in the PDA one) except, as demonstrated in Rogers (1996) by means of computer simulation, when the refractory period is very long, compared with the expected time to speciation. This result of Rogers (1996) holds when only the daughter species have their refractory periods after the birth (see the discussion of the LA-AD model in § 5 of electronic Appendix A).

However, if the parent species have their refractory periods after giving birth as well, then the LA model can be arbitrarily close to a 'completely balanced' (CB) model (see the discussion of the LA-RAD model in § 5 of electronic Appendix A).

Heard (1996) proposed models with rate changes depending on trait changes. His studies suggest that such models can mimic observed imbalance, and a credible explanation (Heard 1996) of this phenomenon is that a species is capable of 'getting stuck'. What was referred to as (quasi-)stabilization in the MR model is in agreement with this 'getting stuck' effect.

It will be shown in § 5 of electronic Appendix A that the models of Losos & Adler (1995) and Heard (1996) (as well as that of Steel & McKenzie (2001)) in fact all belong to the MR family.

Several distinctive features of the general MR family of evolutionary models are seen at this point. The MR family of models contains not only the ERM model but also the PDA model (as well as an entire range from CB to CU models). The MR model exposes the adaptability of 'nomadic' species and prolonged stasis of 'living fossil' species as major factors possibly contributing to the imbalance of phylogenetic trees. Next, the MR family of models is free of arbitrary assumptions, including those on the probability distribution of the trait value changes and on the functional relation between the trait and rate values. Moreover, the MR model has unlimited additional degrees of freedom, as the states of species can carry any amount of additional information, the geographical location being just one of the possible attributes. Also, extinction is taken into account by considering incomplete MR models. In addition, the MR model is functionally simple and requires only three parameters to reproduce the observed imbalance.

The author thanks the referees for useful and stimulating comments and references to Losos & Adler (1995), Harcourt-Brown *et al.* (2001) and Steel & McKenzie (2001).

REFERENCES

Aldous, D. J. 1996 Probability distributions on cladograms. In *Random discrete structures* (ed. D. Aldous & R. Pemantle), pp. 1–18. New York: Springer.

Aldous, D. J. 2001 Stochastic models and descriptive statistics for phylogenetic trees. *Statist. Sci.* **16**, 23–34.

Cunningham, S. 1995 Problems with null models in the study of phylogenetic radiation. *Evolution* **49**, 1292–1294.

Gould, J. S., Raup, D. M., Sepkoski Jr, 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.

Guyer, C. & Slowinski, J. B. 1993 Adaptive radiation and the topology of large phylogenies. *Evolution* **47**, 253–263.

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

Harris, T. E. 1963 *The theory of branching processes*. Berlin: Springer.

Heard, S. B. 1992 Patterns in tree balance among cladistic, phenetic, and randomly generated phylogenetic trees. *Evolution* **46**, 1818–1826.

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

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

Kendall, D. G. 1948 On the generalized 'birth-and-death' process. *Ann. Math. Statist.* **19**, 1–15.

Losos, J. B. & Adler, F. R. 1995 Stumped by trees? A generalized null model for patterns of organismal diversity. *Am. Nat.* **145**, 329–342.

Maddison, W. P. & Slatkin, M. 1991 Null models for the number of evolutionary steps in a character in a phylogenetic tree. *Evolution* **45**, 1184–1197.

Mooers, A. Ø. 1995 Tree balance and tree completeness. *Evolution* **49**, 379–384.

Mooers, A. Ø. & Heard, S. B. 1997 Inferring evolutionary process from phylogenetic tree shape. *Q. Rev. Biol.* **72**, 31–54.

Nee, S. 2001 Inferring speciation rates from phylogenies. *Evolution* **55**, 661–668.

Nee, S., Mooers, A. Ø. & Harvey, P. H. 1992 Tempo and mode of evolution revealed from molecular phylogenies. *Proc. Natl Acad. Sci. USA* **89**, 8322–8326.

Nee, S., Holmes, E. C., May, R. M. & Harvey, P. H. 1994*a* Extinction rates can be estimated from molecular phylogenies. *Phil. Trans. R. Soc. Lond.* B **344**, 77–82.

Nee, S., May, R. M. & Harvey, P. H. 1994*b* The reconstructed evolutionary process. *Phil. Trans. R. Soc. Lond.* B **344**, 305–311.

Nee, S., Holmes, E. C., May, R. M. & Harvey, P. H. 1995 Estimating extinction from molecular phylogenies. In *Extinction rates* (ed. J. H. Lawton & R. M. May), pp. 164–182. Oxford University Press.

Rogers, J. S. 1993 Response of Colless's tree imbalance to number of terminal taxa. *Syst. Biol.* **42**, 102–105.

Rogers, J. S. 1994 Central moments and probability distribution of Colless's coefficient of tree imbalance. *Evolution* **48**, 2026–2036.

Rogers, J. S. 1996 Central moments and probability distributions of three measures of phylogenetic tree imbalance. *Syst. Biol.* **45**, 99–110.

Rosen, D. E. 1978 Vicariant patterns and historical explanation in biogeography. *Syst. Zool.* **27**, 159–188.

Simberloff, D., Heck, K. L., McCoy, E. D. & Connor, E. F. 1981 There have been no statistical tests of cladistic biogeographical hypotheses (with discussion.). In *Vicariance biogeography: a critique* (ed. G. Nelson & D. E. Rosen), pp. 41–93. New York: Columbia University Press.

Slowinski, J. B. 1990 Probabilities of *n*-trees under two models: demonstration that asymmetrical interior nodes are not improbable. *Syst. Zool.* **39**, 89–94.

Slowinski, J. B. & Guyer, C. 1989 Testing the stochasticity of patterns of organismal diversity: an improved null model. *Am. Nat.* **134**, 907–920.

Soltis, P. S., Soltis, D. E., Savolainen, V., Crane, P. R. & Barraclough, T. G. 2002 Rate heterogeneity among lineages of tracheophytes: integration of molecular and fossil data and evidence for molecular living fossils. *Proc. Natl Acad. Sci. USA* **99**, 4430–4435.

Stanley, S. M., Signor III, P. W., Lidgard, S. & Karr, A. F. 1981 Natural clades differ from 'random' clades: simulations and analyses. *Paleobiology* **7**, 115–127.

Steel, M. & McKenzie, A. 2001 Properties of phylogenetic trees generated by Yule-type speciation models. *Math. Biosci.* **170**, 91–112.

Tamas, I., Klasson, L., Canbaeck, B., Naeslund, A. K., Eriksson, A.-S., Wernegreen, J. J., Sandstroem, J. P., Moran, N. A. & Andersson, S. G. E. 2002 50 Million years of genomic stasis in endosymbiotic bacteria. *Nat. Rev. Genet.* **3**, 2376–2379.

Wernegreen, J. J. 2002 Genome evolution in bacterial endosymbionts of insects. *Science* **296**, 850–861.

Yule, G. U. 1924 A mathematical theory of evolution, based on the conclusions of Dr J. C. Willis, F.R.S. *Phil. Trans. R. Soc. Lond.* B **213**, 21–87.

As this paper exceeds the maximum length normally permitted, the author has agreed to contribute to production costs.

Visit http://www.pubs.royalsoc.ac.uk to see an electronic appendix to this paper.