

---

# THE DYNAMICS OF STEM AND CROWN GROUPS

---

A PREPRINT

**Graham E. Budd**

Department of Earth Sciences  
Palaeobiology programme  
Uppsala University, Sweden.  
graham.budd@pal.uu.se

**Richard P. Mann**

Department of Statistics  
School of Mathematics  
University of Leeds, U.K.  
&  
The Alan Turing Institute  
London, U.K.  
r.p.mann@leeds.ac.uk

May 9, 2019

## ABSTRACT

The fossil record of the origins of major groups is of great interests to many biologists, especially when the fossil record apparently conflicts with timings based on molecular clock estimates. Here we model the diversity of “stem” (basal) and “crown” (modern) members of groups as seen in the fossil record, using a “birth-death model”. Under background conditions, the stem group members must diversify rapidly until the modern crown group emerges, at which point their diversity rapidly collapses, followed shortly by their extinction. Mass extinctions can disturb this pattern to create very diverse stem groups such as the dinosaurs and trilobites. Understanding these null-hypothesis patterns is essential for framing ecological and evolutionary explanations for how major groups originate and subsequently evolve.

The fossil record shows many striking patterns [1] that have been addressed by hypotheses such as mass extinctions [2], diversity-dependent diversification [3] and competitive displacement [4, 5]. In particular, the timing and nature of the emergence of modern groups such as animals [6], birds [7, 8] and flowering plants [9] have been of intense interest: here, molecular clock estimates have often been at odds with estimates based solely on the fossil record [9, 10, 11]. Understanding this discrepancy has, however, been hampered by the lack of a model of how both basal (“stem group”) and modern (“crown group”) members [6, 12] of a particular group diversify and go extinct in the fossil record. Quantification of the fossil record, rooted especially in the work of Raup and others [13], allowed explicit models of diversification to be tested, although problems remained about estimation of speciation and extinction rates from the fossil record [14, 15]. Such modelling also failed to take into account the important survivorship and selection biases of the “push of the past” and large clade effects [16], nor the systematic differences that exist between extinct and extant taxa [6, 12]. Any clade can be divided into two components: the last common ancestor of all the living forms and all of its descendants (the “crown group”), and the extinct organisms more closely related to a particular crown group than to any other living group (the “stem group”): together, they make up the “total group” (Fig. 1a). Here we also define the provisional crown group (pCG): the crown group as it would have appeared in the past (e.g., to a Silurian palaeontologist). As basal members of the pCG go extinct through time, the node defining it moves upwards to subtend the next pair of (then) living sister groups, defining a new pCG. When the last member of the stem group goes extinct, the node defining the crown group will be fixed until the present. Nevertheless, the definitive (i.e. modern) crown group will have emerged some time before this, as a subclade of the then pCG, and therefore the definitive crown group will temporarily co-exist with its corresponding stem group. Diversifications can thus be divided into three sequential phases: (1) only the stem group; (2) both stem and crown group; and (3) only the crown group (Fig. 1a).

The origins and fates of crown groups are of particular interest as they comprise modern diversity. However, understanding their origins has been hampered by the lack of any analysis of combined stem and crown group dynamics; the plausibility of models to explain the origin of modern diversity (e.g. Fig. 1b) cannot therefore be easily assessed. An absence of crown group taxa after a time of origin based on a molecular clock estimate could be explained by invoking poor fossil preservation [18]. However, observed stem-group diversity potentially offers a mode of assessing this sort of

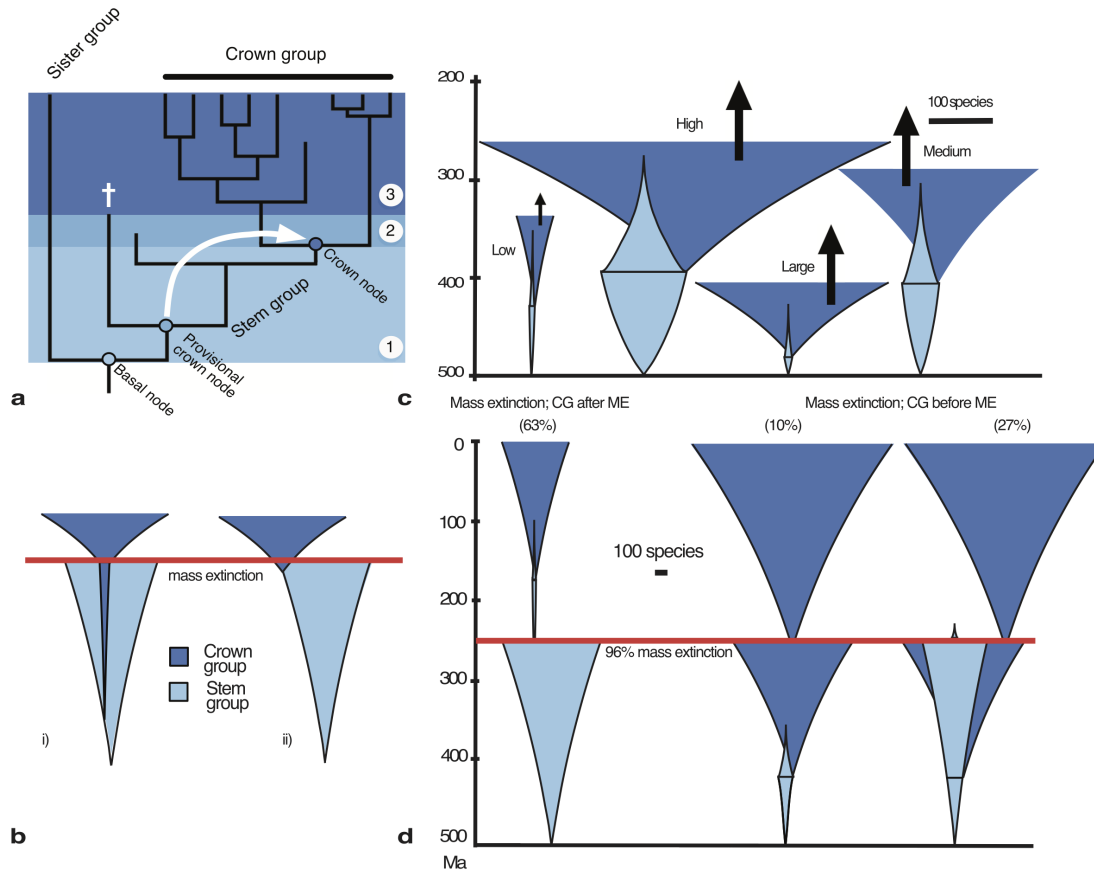


Figure 1: The dynamics of stem and crown groups. **a** Formation of a living crown group is divided into three phases: (1) stem group only; (2) stem and crown group; and (3) crown group only. The definitive crown group is formed as the last stem group member ( $\dagger$ ) goes extinct, and the previous provisional crown group node moves up to the definitive crown group node (white arrow). **b** Two hypotheses to explain the origin of crown groups: i) the “phylogenetic fuse” [17] where the crown group arises early and is cryptic; ii) origin just before a mass extinction. **c** To-scale spindle diagrams of typical stem and crown group diversities for low, middle and high turnover rates (extinction rate = 0.1 (with origination rate 0.1143), 0.5 (origination rate 0.5107) and 1 (origination rate 1.0090) per species per million years, each giving c. 10000 species in the present) and a large clade (here five times larger than expected given its diversification parameters [18]). Note that in all cases (black arrows) the dark blue crown groups continue diversifying to the present. **d** The effect of a large mass extinction (96%) at 250 Ma on the stem and crown groups for a clade with 0.5 background extinction rate that would have been expected to generate 10000 species in the present. The probability of each of the three possible outcomes (crown group after mass extinction; crown group before and stem group dies before; crown group before and stem group reaches mass extinction) are given.

claim if we have a model for relative stem and crown group diversity together. For example: if such a model suggests high crown group diversity when only stem group taxa are observed (e.g. as is the case for birds discussed below), then either the molecular clock estimate for the crown group’s origin is incorrect, or there may be a strong preservation bias between the stem and crown groups that may, on the basis of the model, be further assessed for plausibility.

Here, then, we derive explicit expressions for both stem and crown group diversity and their ratio, using a birth-death model [19]. We first consider how they evolve under homogeneous conditions, and then how mass extinctions perturb the process. In the following, we condition the process on the observation that the clade survives until the present day to form a crown group of at least two species. We take as a model example a total group that emerged 500 Ma, and whose crown group emerged c. 410 Ma; and with diversification parameters speciation  $\lambda = 0.5107$  and extinction  $\mu = 0.5$  (both per million years) (c.f. fig. 1 of ref. [16]). These numbers would on average generate 10,000 species in the Recent, assuming no mass extinctions. We also consider similarly-sized clades generated with high turnover ( $\mu = 1, \lambda = 1.0090$ ) and low turnover ( $\mu = 0.1, \lambda = 0.1143$ ), and how unusually large clades [16, 20] (given their

diversification parameters) behave. We then introduce mass extinctions [21] of various sizes. In all cases we calculate for 1 million year intervals the expected size of the stem group; the probability that it has gone extinct; the expected proportion of total diversity contained within the stem at any given time (averaged over all clades with these parameters); the probability that the stem contains a certain proportion of diversity, and the expected absolute diversity in the stem (Fig 2). The spindle diagrams of Fig. 1c, d show typical to-scale shapes and sizes of the stem and crown groups under these conditions: note that because of the high stochasticity of the process, a wide range of less-likely outcomes are also possible [16].

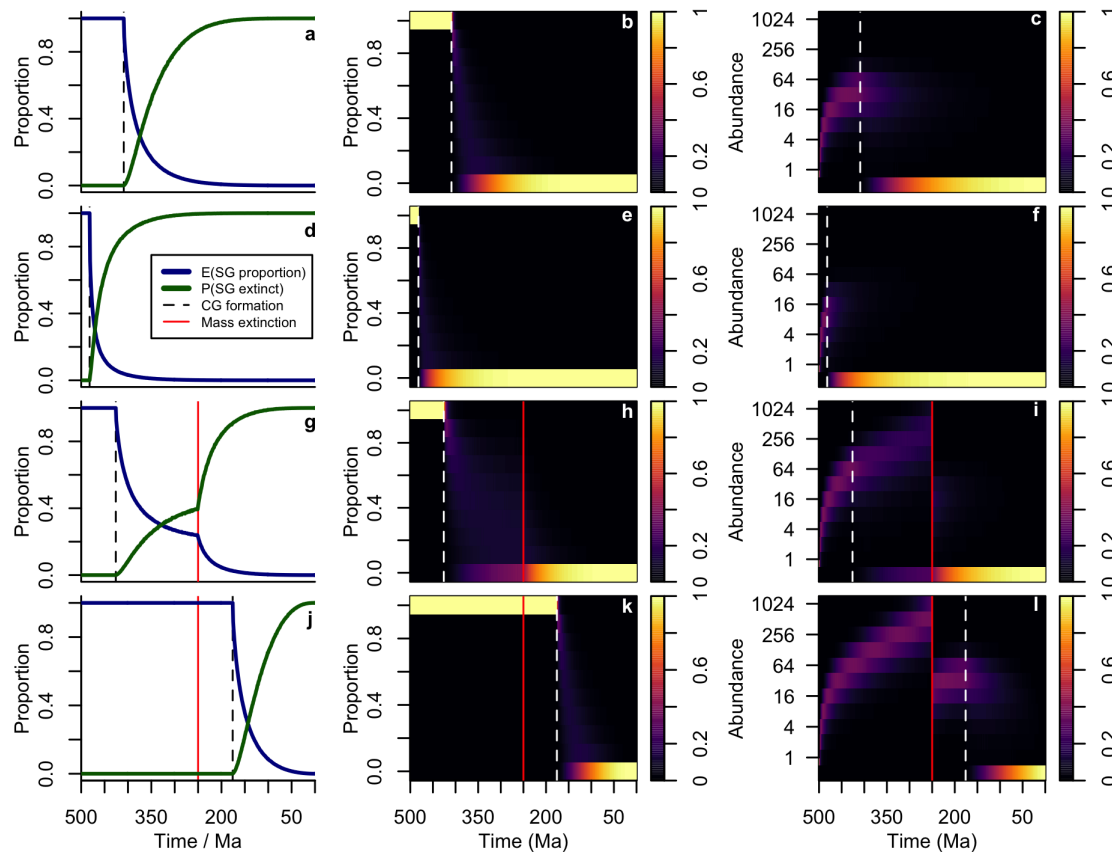


Figure 2: Stem group extinction and diversity. First column: expected total diversity contained in the stem (blue line), probability that stem group has gone extinct (green) line, selected time of origin of crown group (black dashed line) and time of mass extinction (vertical red line); second column; probability heat map of proportion of diversity being contained in the stem group; third column, probability heat map of absolute abundance of stem group through time. In columns 2 and 3, red lines represent a mass extinction and the dashed white line the time of origin of crown group. **a-c**: the “typical” clade of expected size 10000 in the present with speciation rate = 0.5107 and extinction rate = 0.5; **d-f**: a “large clade”, five times larger than expected given its diversification parameters; **g-i**: the typical clade but with a 96% mass extinction at 250 Ma when the crown group forms before the mass extinction (37% of the time); and **j-l**: as **g-i**, but when the crown group forms after the mass extinction (63% of the time). E(SG proportion): expected proportion of total diversity taken up by stem group; P(SG extinct): proportion of trees in which the stem group has gone extinct. The dashed line represents the time of formation of the crown group (here selected as the mean time from eqn 11, 12 below) and the red line (where present) the time of a mass extinction.

## Results

In our model example, after the emergence of the crown group, the average proportion of diversity contained within the stem group drops exponentially; the stem group makes up half of expected total diversity only c. 15 Myrs after the origin of the crown group. By c. 80 Myrs after, it is expected to make up only 10% of total group diversity. In addition, the probability of the stem group having gone entirely extinct climbs sharply in the same interval: to 50% after 60 Myrs and 75% after 100 Myrs. With high turnover, the stem-group grows faster initially, lasts longer, and generates a

larger crown group; however, it too declines extremely quickly once the crown group is established. The converse of all these is true for the low turnover example, except that here too the stem group declines quickly after the crown group origin. These quantitative results illustrate an important insight: crown groups and the stem-group before the crown group emerge rapidly as a necessary condition for survival to the present; and stem groups after the emergence of the crown group cease expansion and rapidly decline as the necessary condition for their extinction. The net effect is that a pCG always diversifies rapidly. The rapidity of the transition from a stem-dominated to a crown-dominated regime is sharpened when the crown group emerges more quickly than expected, as would be the case for an unusually large clade [16, 20]. One would expect the crown group of a clade (such as the arthropods) that ends up perhaps 5 times larger than expected over the same 500 Myrs to emerge at only c. 20 Myrs after the total group (Fig 1b), and for its stem to dwindle very quickly. Provisional crown groups quickly become converted to the definitive crown group, and the crown group node (Fig 1A) then becomes stable. In our example, the definitive crown group forms after an average of 90 Myrs and the stem dies an average of 110 Myrs later, implying that the pCG stops changing from this point and that the identity of the crown group node remains consistent until the present. Such stability explains why the fossil record is dominated by long blocks of time with a very similar biota [2] (e.g. the “Age of Reptiles”). Once a pCG has sufficiently diversified, it becomes very unlikely to go extinct by stochastic variation alone. Such pCGs and their shared characteristics thus become stable features that are disrupted only by the largest mass extinctions.

Long-lived groups have been through one or more mass extinctions. The effect of a very large (96%) mass extinction at 250 Ma on our typical clade is shown in Figs 1d and 2c, d, with the three possibilities of crown group emergence after the mass extinction (63%); crown group emergence before the mass extinction with stem group survival until the mass extinction (27%); or crown group emergence before the mass extinction with the stem group becoming extinct before the mass extinction (10%) being illustrated. The probability of the crown group emerging after a mass extinction is only high when the absolute number of surviving species is small (Fig 3b, c). For our example clade, the tipping point mass extinction size (when the crown group becomes more likely to emerge after than before the extinction) is as much as 93%. Well-established (i.e. relatively long-lived) provisional crown groups, like crown groups themselves, are also robust even to large mass extinctions; the loss of famous clades such as trilobites seems to be the result of successive biotic crises, not a single event [22].

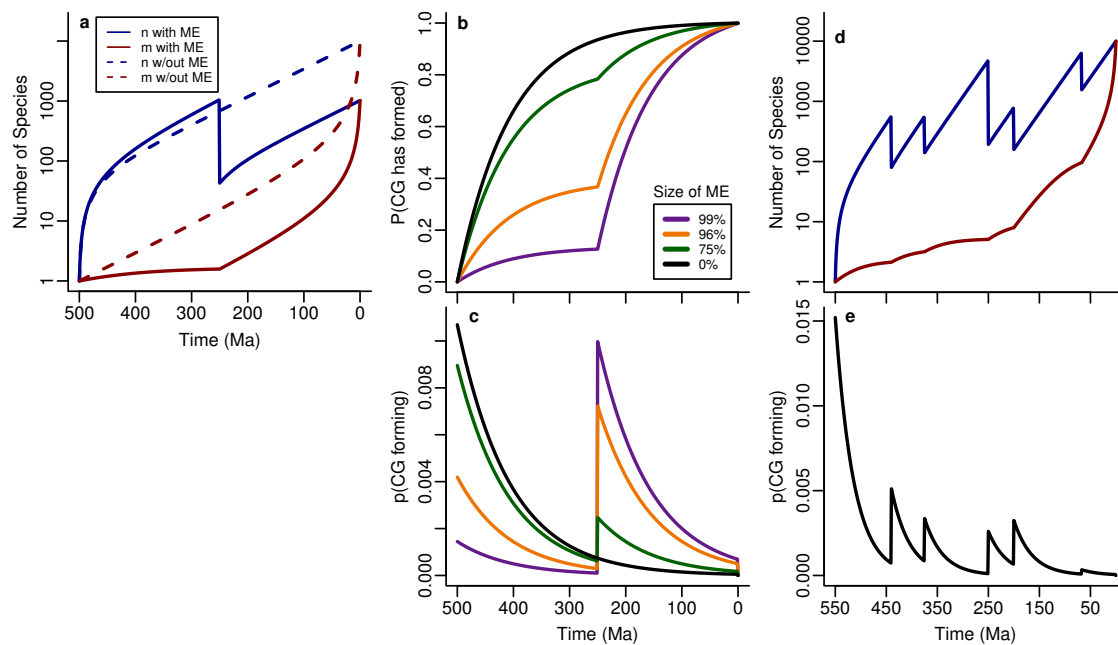


Figure 3: The effects of single and multiple mass extinctions. **a** Species (blue) and lineage (red) plots for a typical diversification that would normally generate 10,000 species in the present over 500 Myrs with background extinction rate 0.5 and speciation rate c. 0.5107 (dotted lines) but is affected by a 96% mass extinction at 250 Ma (note slightly higher initial push of the past 18). **b, c** Cumulative and non-cumulative probability distribution curves for probability that crown group has formed for mass extinctions of various magnitudes; note sharp increase in all cases directly after the mass extinction, and that only the largest mass extinctions have a significant effect. **d, e** The effect over Phanerozoic time (c. 550 Myrs) of the canonical “big five” mass extinctions [23] on a typical clade originating at the beginning of the Cambrian on **d** species and lineage counts and **e** probability of crown group forming.

In Fig. 3b, c we show the fates of both species and lineages counts through time as they pass through a mass extinction of varying size (for that particular clade) and the probability of the emergence of the crown group against time. The bifurcating effect of mass extinctions on crown group origins implies that the least likely time for a crown group to emerge is just before a mass extinction (Fig. 3c). For example, if crown Aves really evolved at around 70 Ma8 and *Vegavis* [24] is an anseriform, then at least five crown group bird lineages independently crossed the K/Pg boundary [24] and at the same time had an origin very close to it in time. Our modelling shows this to be an unlikely possibility (c.f. ref. [25]), and that the controversial “explosive” post-Cretaceous model [10] of crown group birds is more likely.

It is also possible to calculate directly the effect of multiple mass extinctions [21], such as the canonical “big five” of the Phanerozoic [23]. For a set of such extinctions across Phanerozoic time (Fig. 3 d, e) it can be seen that the probability of the crown group of a Cambrian total group emerging is sharply higher in periods of time just after mass extinctions, apart from the continued high likelihood of a Cambrian origin. The suggestive resemblance of Fig 3e to plots of Phanerozoic high-level taxonomic origination rates [3, 14] may not be coincidental, with higher level taxa such as families that appear to preferentially appear after mass extinctions corresponding partly to crown groups.

## Discussion

Our results reveal much about the dynamics of stem and crown groups and have an important bearing on how we view the fossil record. In particular, as the crown group forms, the stem group quickly collapses into first obscurity and then extinction unless the total group was affected by a very large mass extinction. Indeed, this rapid phase transition from stage 1 to stage 3 of the diversification process might be mistaken for a mass extinction itself, such as has been suggested for the end of the Ediacaran period [26]. Qualitative models [17] wherein crown groups form “silently” and remain as low diversity components of the background while the stem group diversifies (Fig. 1b) can be seen to be unlikely under the model; rather, the reverse is always true. This is particularly important in clades where molecular clocks suggest a deep origin of the crown group during an interval of time when the stem group is seen to be diversifying in the fossil record, with classic examples being animals, angiosperms, birds and mammals [27, 9, 28, 18, 11] (Fig. 1b). In common with other recent analyses [9, 16, 27], this suggests that systematic errors in molecular clock estimates, especially for large clades, are creating illusory deep roots that are profoundly at odds with both modelling of diversification and observed patterns in the fossil record.

The famous debates in the 1970s in the pages of *Paleobiology* and elsewhere [29] were essentially between viewing the fossil record as an archive of discernable evolutionary processes (such as diversity-dependent diversification [3]) or as the result of inevitable stochastic patterns that could be captured with simple statistical models [30]. We have previously extended the latter view by modelling heterogeneity that results from survivorship and selection bias [16]. Here we show that other types of heterogeneity in the record such as the rapid takeover of diversity by the crown group and preferential crown group emergence after large mass extinctions also emerge from imposing retrospective structure on a homogeneous process and should therefore also be part of any null hypothesis. A homogeneous model is not of course, intended to capture the many true heterogeneities that exist in the evolutionary process: it is a model, not a description. Nevertheless, it shows both that strong patterns of heterogeneity emerge even under homogeneous background conditions when simple conditions such as survival to the present are imposed, and also that hypotheses that strongly transgress against the general predictions of the model should not simply be accepted without scrutiny, especially when such transgressions are themselves generalised (e.g. the “phylogenetic fuse” [17] concept).

## Mathematical methods

The fundamental mathematical results concerning the distribution of species and lineage numbers in a birth-death process are given by Nee et al. [19], and have been much elaborated and applied subsequently, especially for the reconstructed process (i.e. models for lineages leading to extant species) [21, 31, 32]. Here we use those results to derive properties of the stem and crown groups, both in homogeneous processes and those experiencing mass extinctions that occur at singular points in time, with references to related approaches for the reconstructed process where appropriate. Our notation follows that used in Budd and Mann [16], but we strive to make clear the connection to the notation of Nee et al. [19] and others.

### Glossary of mathematical terminology

- $\lambda$ : the speciation rate of a birth-death process (BDP), assumed to be constant.
- $\mu$ : the extinction rate of a BDP, assumed to be constant.
- $T$ : the time from the origin of a BDP to the present day. Equivalently the age of the total group.

- $n_{t,t'}$ : the number of species alive at time  $t'$  in a BDP that originated at time  $t$ .
- $m_{t,t'}$ : the number of species alive at time  $t$  with descendants at time  $t'$ , in a BDP that originated at time zero.
- $a_{t,t'}$ : parameter controlling the distribution of abundance such that  $(1 - a_{t,t'})$  is the “success” parameter in the geometric distribution of  $n_{t,t'}$  (see equation 1).
- $S_{t,t'}$ : the probability that a BDP originating at time  $t$  has any descendants at time  $t'$ .
- $f_i$ : the fraction of species that survive mass extinction event  $i$ .
- $n_t^{CG}$ : the number of species at time  $t$  in the crown group of a BDP that originated at time zero.
- $n_t^{SG}$ : the number of species at time  $t$  in the stem group of a BDP that originated at time zero.

## The birth-death process

The number of living species,  $n_{t,t'}$  at time  $t'$  in a process that originates at time  $t$ , conditioned on the survival of the process to  $t'$  is geometrically distributed with success parameter  $1 - a_{t,t'}$ :

$$\begin{aligned} P(n_{t,t'}) &= a_{t,t'}^{(n_{t,t'}-1)} (1 - a_{t,t'}) \\ &\equiv G(n_{t,t'} - 1, 1 - a_{t,t'}) \end{aligned} \quad (1)$$

The parameter  $a_{t,t'}$  is analogous to the  $u_t$  of Nee et al. [19], and also corresponds to the lower tail distribution of the coalescent point process  $P(H < T)$  described by Lambert and Stadler [21], where  $H$  is the coalescence time and  $T$  is as above. In a homogeneous process without mass extinctions,  $a_{t,t'}$  is a function of the background speciation rate  $\lambda$  and extinction rate  $\mu$ :

$$a_{t,t'} = \frac{\lambda(1 - \exp(-(\lambda - \mu)(t' - t)))}{\lambda - \mu \exp(-(\lambda - \mu)(t' - t))} \quad (2)$$

Nee et al.[19] showed that the number of species alive at time  $t$  with descendants at time  $t'$ ,  $m_{t,t'}$  is also geometrically distributed (note Nee et al. [19] use the notation  $\eta$  rather than  $m$ ):

$$P(m_{t,t'}) = G(m_{t,t'} - 1, (1 - a_{0,t} S_{0,t'} / S_{0,t})), \quad (3)$$

where  $S_{t,t'}$  is the probability for a birth-death process originating with one species at time  $t$  to survive to time  $t'$  (analogous to the  $P(t, t')$  of Nee et al.[19]). In a homogeneous process this survival probability is given as:

$$S_{t,t'} = \frac{\lambda - \mu}{\lambda - \mu \exp(-(\lambda - \mu)(t' - t))} \quad (4)$$

## Mass extinctions

As shown by Nee et al. [19], the distributions of  $n_{t,t'}$  and  $m_{t,t'}$  in terms of  $a_{t,t'}$  and  $S_{t,t'}$  continue to hold when the process is not homogeneous, for example when mass extinctions occur. However, the forms of  $a_{t,t'}$  and  $S_{t,t'}$  must change to account for this heterogeneity. The parameter controlling species diversity,  $a_{t,t'}$ , can be given more generally in terms of time dependent  $\lambda(s)$  and  $\mu(s)$  as:

$$a_{t,t'} = 1 - S_{t,t'} \exp\left(-\int_t^{t'} \lambda(s) - \mu(s) ds\right). \quad (5)$$

In the case of a set of mass extinctions occurring instantaneously at times between  $t$  and  $t'$ , with survival fractions  $f_1, f_2, \dots, f_N$ , in an otherwise homogeneous process, this reduces to:

$$a_{t,t'} = 1 - S_{t,t'} \exp(-(\lambda - \mu)(t' - t) + \sum_{i=1}^N \ln(f_i)). \quad (6)$$

What remains is to determine the survival probability  $S_{t,t'}$  of the process through the mass extinctions. First we consider survival through one mass extinction, and then identify a recursive process for calculating the survival probability through multiple mass extinctions. Consider a birth death process starting with one species at time  $t$ . If each species has a instantaneous and independent probability  $f$  to survive a mass extinction at time  $t_m$ , what is the probability that the process has descendants at time  $t'$ , beyond the mass extinction?

Firstly, the distribution of the number of descendants it has at time  $t_m$ , from equation 1, is:

$$P(n_{t,t_m}) = S_{t,t_m} a_{t,t_m}^{(n_{t,t_m}-1)} (1 - a_{t,t_m}) \quad (7)$$



The probability that any one of these descendants at time  $t_m$  will leave a descendent at time  $t'$  is given by the product of the (independent) probability that it survives the mass extinction ( $f$ ) and the probability that its lineage survives another time  $t' - t_m$ . Therefore the probability that at least one will survive is:

$$P(n_{t,t'} > 0 | n_{t,t_m}) = 1 - (1 - fS_{t_m,t'})_{t,t_m}^n, \quad (8)$$

To obtain the probability  $S_{t,t'}$  that any species at time  $t$  will have some descendants at time  $t'$  we must marginalize over the unknown value of  $n_{t,t_m}$ :

$$\begin{aligned} S_{t,t'} &= S_{t,t_m} \sum_{n_{t,t_m}=1}^{\infty} P(n_{t,t'} > 0 | n_{t,t_m})P(n_{t,t_m}) \\ &= \frac{S_{t,t_m} (1 - (1 - fS_{t_m,t'})(1 - a_{t,t_m}))}{(1 - (1 - fS_{t_m,t'})a_{t,t_m})} \end{aligned} \quad (9)$$

In the case of one mass extinction,  $a_{t,t_m}$ ,  $S_{t,t_m}$  and  $S_{t_m,t'}$  are all given by equations 2 and 4, since these periods contain no further heterogeneities. However, in the case of multiple mass extinctions, equation 9 can be applied recursively along with equation 6 to obtain these values, until all mass extinctions have been incorporated.

Using these adapted expressions for  $a_{t,t'}$  and  $S_{t,t'}$ , the geometric distributions specified in equations 1 and 3 continue to hold. We note that Lambert and Stadler [21] considered the reconstructed evolutionary process as a coalescent point process (CPP), and concisely derived a full distribution of trees through an arbitrary set of mass extinctions, showing that the process remains a CPP. This prior result thus provides an alternative derivation for our results concerning the number of lineages that survive to the present,  $m_{t,T}$ , as these constitute the reconstructed phylogeny. However, the reconstructed process by definition excludes the extinct stem lineages.

### When does the crown group form?

The probability that the crown group has formed by time  $t$  is the probability that the number of species alive at time  $t$  that have descendants in the present,  $m_{t,T}$  is greater than one. We take a crown group here to consist of at least two species, so that both the crown and stem group can be defined. Since  $m_{T,T} > 1$  (i.e. we observe that the crown group does form before the present), this implies:

$$P(\text{CG has formed by } t | \text{CG has formed by } T) = P(m_{t,T} > 1 | m_{T,T} > 1) \quad (10)$$

Using equation 3 we have:

$$\begin{aligned} P(m_{t,T} > 1 | m_{T,T} > 1) &= \frac{P(m_{t,T} > 1 | m_{T,T} > 0)}{P(m_{T,T} > 1 | m_{T,T} > 0)} \\ &= \frac{(a_{0,t}S_{0,T})}{(a_{0,T}S_{0,t})} \end{aligned} \quad (11)$$

We can further condition on a known limit to the time of the crown group. For example, if we wish to consider cases where the crown group emerges before the time of a mass extinction,  $t_m$ , the following the above gives an updated expression:

$$P(\text{CG has formed by } t | \text{CG has formed by } t_m) = \frac{(a_{0,t}S_{0,t_m})}{(a_{0,t_m}S_{0,t})} \quad (12)$$

This result can also straightforwardly be derived by the CPP approach of Lambert and Stadler [21] since the crown group node subtends all extant species of a clade.

### Dynamics of stem and crown group abundance

In this section we derive the new mathematics required to model stem and crown group abundance together. We assume that the process starts at time  $t = 0$  and that the time at which the crown group emerges,  $t_c$  is known, and derive expressions for abundance in the stem and crown groups conditioned on this.

### Before the crown group forms

Before the crown group has formed, all diversity is stem diversity. We can calculate the distribution of this diversity,  $n_t^{SG}$ , by conditioning on  $m_{t,T} = 1$ , i.e. that the crown group has not yet formed, revealing a negative binomial distribution:

$$\begin{aligned} P(n_{0,t} | m_{t,T} = 1) &= \frac{P(n_{0,t})P(m_{t,T} = 1 | n_{0,t})}{P(m_{t,T} = 1)} \\ &= n_{0,t}(1 - a_{0,t}(1 - S_{t,T}))^2(a_{0,t}(1 - S_{t,T}))^{(n_{0,t}-1)} \\ &\equiv \text{NB}(n_{0,t} - 1, 2, (1 - a_{0,t}(1 - S_{t,T}))) \end{aligned} \quad (13)$$

### After the crown group forms

*The crown group:* The crown group is defined by a speciation at time  $t_c$ , creating exactly two species that will give rise to descendants in the present. This implies two independent BD processes that will survive to time  $T$ , starting at time  $t_c$ , thus implying for each that  $m_{t_c,T} > 0$ . This implies [19] that the abundance for each of these processes is the sum of two non-identically distributed geometric random variables,  $X$  and  $Y$ , with:

$$\begin{aligned} P(X) &= \text{G}(X - 1, 1 - a_{t_c,t}), X > 0 \\ P(Y) &= \text{G}(Y, 1 - a_{t_c,t}(1 - S_{t,T})), Y \geq 0 \end{aligned} \quad (14)$$

Since the crown group consists of two such independent BD processes, we can use the fact that the sum of two i.i.d. geometric random variables follows a negative binomial distribution to express the crown group diversity at time  $t$ :

$$n_t^{CG} = W + Z \quad \forall t > t_c \quad (15)$$

with

$$\begin{aligned} P(W) &= \text{NB}(W - 2, 2, (1 - a_{t_c,t}), W \geq 2 \\ P(Z) &= \text{NB}(Z, 2, (1 - a_{t_c,t}(1 - S_{t,T})), Z \geq 0 \end{aligned} \quad (16)$$

*The stem group:* Let  $m_{t_c,t}$  be the number of species just before the crown group speciation event that give rise to descendants at time  $t$ . We know that  $m_{t_c,T} = 1$ , since one species will form the base of the crown group, while the others will give rise to the stem, which is extinct at time  $T$ . Conditioning on this, we have:

$$\begin{aligned} P(m_{t_c,t} | m_{t_c,T} = 1) &\propto P(m_{t,t'})P(m_{t,T} = 1 | m_{t,t'}) \\ &\propto \binom{m_{t,t'}}{m_{t,t'} - 1} \left(1 - a_{0,t} \frac{S_{0,t'}}{S_{0,t}}\right) \left(a_{0,t} \frac{S_{0,t'}}{S_{0,t}}\right)^{(m_{t,t'}-1)} \left(1 - \frac{S_{t,T}}{S_{t_c,T}}\right)^{(m_{t,t'}-1)} \left(\frac{S_{t,T}}{S_{t_c,T}}\right) \end{aligned} \quad (17)$$

With correct normalisation this reveals a negative binomial distribution:

$$P(m_{t_c,t} | m_{t_c,T} = 1) = \text{NB} \left( m_{t_c,t} - 1, 2, 1 - a_{0,t} \left( \frac{S_{0,t}}{S_{0,t_c}} \right) \left( 1 - \frac{S_{t,T}}{S_{t_c,T}} \right) \right) \quad (18)$$

We can identify the quantity  $m_{t_c,t} - 1$  as the number of species at time  $t_c$  prior to the crown group speciation that have descendants in the stem at time  $t$ , since one species at time  $t_c$  only has descendants in the crown group. The total stem diversity at time  $t$  is therefore the sum of  $m_{t_c,t} - 1$  independent BD processes that are conditioned on surviving for a duration  $t - t_c$  but also on becoming extinct before time  $T$  (by definition the stem is extinct at time  $T$ ). First we consider the dynamics of one such BD process. From equation 1, conditioning only on survival to time  $t$  from a start of  $t_c$ , we have:

$$P(n_{t_c,t}) = a_{t_c,t}^{(n_{t_c,t}-1)}(1 - a_{t_c,t}) \quad (19)$$

Now we also condition on the process not surviving to time  $T$ :

$$\begin{aligned} P(n_{t_c,t} | \text{Extinct at } T) &\propto P(n_{t_c,t})P(\text{Extinct at } T | n_{t_c,t}) \\ &\propto a_{t_c,t}^{(n_{t_c,t}-1)}(1 - a_{t_c,t})(1 - S_{t,T})^{n_{t_c,t}} \end{aligned} \quad (20)$$

With normalisation this reveals another geometric distribution:

$$P(n_{t_c,t} | \text{Extinct at } T) = \text{G}(n_{t_c,t}, (1 - a_{t_c,t}(1 - S_{t,T}))) \quad (21)$$

We can now see the total stem abundance at time  $t$  as the sum of  $m_{t_c,t} - 1$  i.i.d. geometrically-distributed random variables, where  $m_{t_c,t} - 1$  is itself composed of the sum of two geometrically distributed random variables (by virtue



of its own negative binomial distribution). Marginalising over possible values of  $m_{t_c,t} - 1$ , this gives a full marginal distribution for the stem diversity  $n_t^{SG}$  for times  $t > t_c$ :

$$\begin{aligned}P(n_t^{SG} = 0) &= r^2 \\P(n_t^{SG} = 1) &= 2(1-r)rG(0, rq) \\P(n_t^{SG} = x) &= 2(1-r)rG(x-1, rq) + (1-r)^2\text{NB}(x-2, 2, rq)\forall x > 1\end{aligned}\tag{22}$$

where  $q = 1 - a_{t_c,t}(1 - S_{t,T})$  and  $r = 1 - a_{0,t}(S_{0,t}/S_{0,t_c})(1 - (S_{t,T}/S_{t_c,t}))$ , and the geometric and negative binomial mass functions G and NB are defined as above. A direct corollary of this result is that the probability that the stem is extinct at time  $t$  is  $r^2$ .

## References

- [1] M J Benton. Diversification and extinction in the history of life. *Science*, 268(5207):52–58, 1995.
- [2] D M Raup and J J Sepkoski. Mass extinctions in the marine fossil record. *Science*, 215(4539):1501–1503, 1982.
- [3] J John Sepkoski. Rates of speciation in the fossil record. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 353(1366):315–326, 1998.
- [4] S J Gould and C B Calloway. Clams and brachiopods - ships that pass in the night. *Paleobiology*, 6(4):383–396, 1980.
- [5] M J Benton. Progress and competition in macroevolution. *Biological Reviews*, 62(3):305–338, 1987.
- [6] G E Budd and S Jensen. A critical reappraisal of the fossil record of the bilaterian phyla. *Biological Reviews of the Cambridge Philosophical Society*, 75(2):253–295, 2000.
- [7] W Jetz, G H Thomas, J B Joy, K Hartmann, and A O Mooers. The global diversity of birds in space and time. *Nature*, 491(7424):444–448, 2012.
- [8] R O Prum, J S Berv, A Dornburg, D J Field, J P Townsend, E M Lemmon, and A R Lemmon. A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. *Nature*, 526(7574):569–U247, 2015.
- [9] J M Beaulieu, B C O’Meara, P Crane, and M J Donoghue. Heterogeneous Rates of Molecular Evolution and Diversification Could Explain the Triassic Age Estimate for Angiosperms. *Systematic Biology*, 64(5):869–878, 2015.
- [10] A Feduccia. ‘Big bang’ for tertiary birds? *Trends in Ecology & Evolution*, 18(4):172–176, 2003.
- [11] Mike Foote, John P. Hunter, Christine M. Janis, and J. John Sepkoski. Evolutionary and preservational constraints on origins of biologic groups: Divergence times of eutherian mammals. *Science*, 1999.
- [12] AJ Craske and RPS Jefferies. A new mitrate from the upper ordovician of norway, and a new approach to subdividing a plesion. *Palaeontology*, 32(1):69–99, 1989.
- [13] J John Sepkoski. A kinetic model of phanerozoic taxonomic diversity i. analysis of marine orders. *Paleobiology*, 4(3):223–251, 1978.
- [14] J. Alroy. Dynamics of origination and extinction in the marine fossil record. *Proceedings of the National Academy of Sciences*, 105:11536–11542, 2008.
- [15] Tanja Stadler, Alexandra Gavryushkina, Rachel CM Warnock, Alexei J Drummond, and Tracy A Heath. The fossilized birth-death model for the analysis of stratigraphic range data under different speciation modes. *Journal of theoretical biology*, 447:41–55, 2018.
- [16] G E Budd and R P Mann. History is written by the victors: The effect of the push of the past on the fossil record. *Evolution*, 72(11):2276–2291, 2018.
- [17] Alan Cooper and Richard Fortey. Evolutionary explosions and the phylogenetic fuse. *Trends in Ecology and Evolution*, 13(4):151–156, 1998.
- [18] Mario Dos Reis, Yuttapong Thawornwattana, Konstantinos Angelis, Maximilian J Telford, Philip CJ Donoghue, and Ziheng Yang. Uncertainty in the timing of origin of animals and the limits of precision in molecular timescales. *Current Biology*, 25(22):2939–2950, 2015.
- [19] S Nee, R M May, and P H Harvey. The reconstructed evolutionary process. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, 344(1309):305–311, 1994.
- [20] M W Pennell, B A J Sarver, and L J Harmon. Trees of Unusual Size: Biased Inference of Early Bursts from Large Molecular Phylogenies. *Plos One*, 7(9), 2012.

- [21] Amaury Lambert and Tanja Stadler. Birth-death models and coalescent point processes: The shape and probability of reconstructed phylogenies. *Theoretical Population Biology*, 2013.
- [22] Mark Bell. Fossil focus: trilobites. *Palaeontol. Onl*, 3:1–9, 2013.
- [23] Anthony D Barnosky, Nicholas Matzke, Susumu Tomiya, Guinevere O U Wogan, Brian Swartz, Tiago B Quental, Charles Marshall, Jenny L McGuire, Emily L Lindsey, Kaitlin C Maguire, Ben Mersey, and Elizabeth A Ferrer. Has the Earth’s sixth mass extinction already arrived? *Nature*, 471:51, mar 2011.
- [24] J A Clarke, C P Tambussi, J I Noriega, G M Erickson, and R A Ketcham. Definitive fossil evidence for the extant avian radiation in the Cretaceous (vol 433, pg 305, 2005). *Nature*, 444(7120):780, 2006.
- [25] Jingmai O’Connor, Luis M. Chiappe, and Alyssa Bell. Pre-modern Birds: Avian Divergences in the Mesozoic. In *Living Dinosaurs: The Evolutionary History of Modern Birds*. 2011.
- [26] Simon A F Darroch, Erik A Sperling, Thomas H Boag, Rachel A Racicot, Sara J Mason, Alex S Morgan, Sarah Tweedt, Paul Myrow, David T Johnston, and Douglas H Erwin. Biotic replacement and mass extinction of the Ediacara biota. In *Proc. R. Soc. B*, volume 282, page 20151003. The Royal Society, 2015.
- [27] Jacob S. Berv and Daniel J. Field. Genomic Signature of an Avian Lilliput Effect across the K-Pg Extinction. *Systematic Biology*, 2018.
- [28] Mario Coiro, James A Doyle, and Jason Hilton. How deep is the conflict between molecular and fossil evidence on the age of angiosperms? *New Phytologist*, 2019.
- [29] D Sepkoski. *Rereading the fossil record: the growth of paleobiology as an evolutionary discipline*. University of Chicago Press, Chicago, 2012.
- [30] D M Raup, S J Gould, T J M Schopf, and D S Simberloff. Stochastic-models of phylogeny and evolution of diversity. *Journal of Geology*, 81(5):525–542, 1973.
- [31] Tanja Stadler. Inferring speciation and extinction processes from extant species data. *Proceedings of the National Academy of Sciences*, 108(39):16145–16146, 2011.
- [32] T A Heath, J P Huelsenbeck, and T Stadler. The fossilized birth-death process for coherent calibration of divergence-time estimates. *Proceedings of the National Academy of Sciences of the United States of America*, 111(29):E2957–E2966, 2014.