Is Diversification Limited? - A birth-death model comparison

Disa Hansson

Is Diversification Limited? - A birth-death model comparison

Disa Hansson[*]

Juni 2011

Abstract

The aim of this thesis is to study how the two events of evolution, speciation and extinction, has changed in 32 bird phylogenies. The main question to answer is if the diversification rate (speciation rate using a modified birth-death process called the reconstructed process, describing the process of a phylogeny where the extinct species has been removed. Six different diversification scenarios, two with constant diversification rate and four with decreasing diversification rate, have been tested on the different phylogenies. A comparison of how well the models perform under the assumption that all extant (now living) species is represented in the phylogeny, i.e. complete sampling, compared to the assumption that there only is a subsample of all extant species in the phylogeny, i.e. random sampling, is also done. The result in this thesis indicates that the diversification rate is decreasing because of a declining speciation rate.

[*]Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail: hansson.disa@gmail.com . Supervisor: Sebastian Höhna.

Acknowledgements

Contents

1 Introduction

Evolution is the biological process that alters life forms on earth. The evolutionary theory tries to describe the common ancestry of all present living species. Due to natural selection species adapt to their environment and under different circumstances, such as geographic separation, different population of the same species can become independent species. The event when two species originate from a single ancestral species is called a speciation event. Furthermore, species compete with each other for resources such as food and space. This competition may lead to that species go extinct.

In macroevolution (evolution on species level) studies are often interested in the patterns of evolution, such as diversification dynamics (Morlon, Potts, & Plotkin, 2010; Lieberman, 2001). Many studies use constant speciation and extinction rates (Lieberman, 2001), although this assumption is not realistic. Studies have for example shown rapid radiation in lizard's early history (Jackman, Larson, De Queiroz, & Losos, 1999). Five mass extinction periods have been observed from the fossil record, extinction periods were followed by periods of rapid radiations (Seposki, 1998). This indicates that rates have not been constant through time and that there is a need to use varying rates (Nee, May, & Harvey, 1994).

Mathematically, a birth-death process models the process of speciation and extinction. A more detailed description is given in Section 2. In this thesis we are interested in modelling the birth-death process and inferring speciation and extinction rates for different sets of species. Speciation and extinction rates can be inferred from molecular phylogenies (Nee, Holmes, May, & Harvey, 1994).

The data in the study of evolution could either be organisms phenotype (colour, shape) or as in molecular phylogenetics their genotype (DNA, RNA). A phylogenetic tree represent the process of evolution by a binary tree, where the ancestral node in the tree represents the most recent common ancestor (MRCA) of the two descendant species and the time since the speciation event is represented by the branch length. An example is illustrated in figure 1.
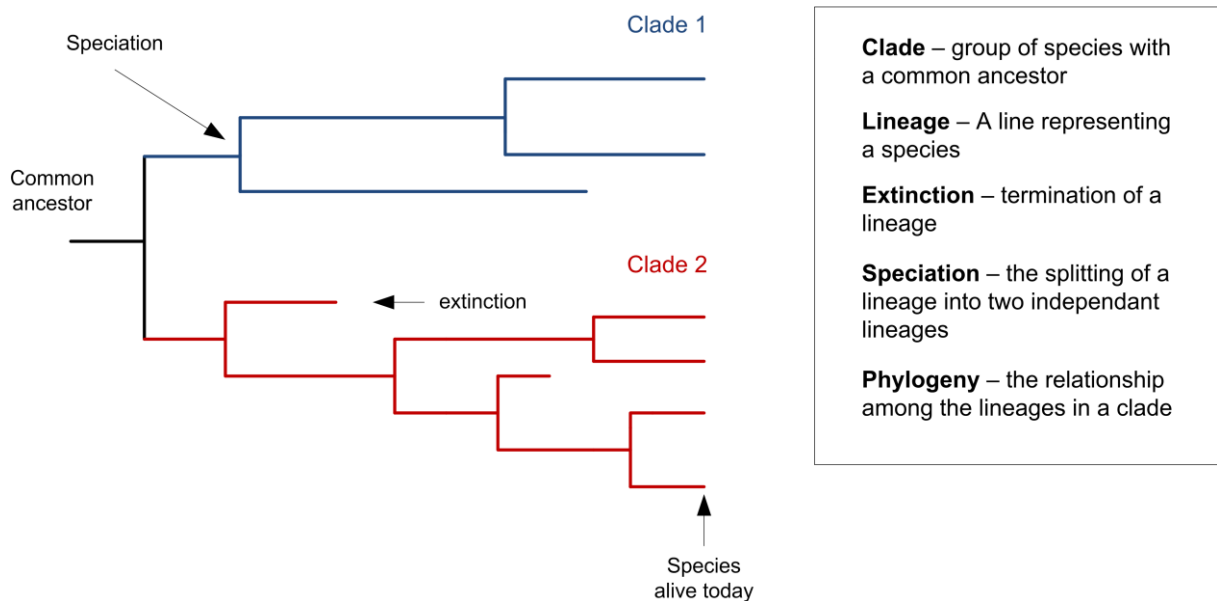
Figure 1 Example of phylogenetic tree and explanations.

In this thesis we will study in what way the two events of evolution, speciation and extinction, have operated through time in 32 bird phylogenies (754 species). Primarily we will try to answer if the diversification rate (=birth rate – death rate) is limited, which could indicate that there is a limit to diversity. The reason for a limit to diversity could be that species now had time to adapt to the available spaces and the environment is becoming saturated, thus speciation events is becoming less frequent. Another reason could be that extinction is becoming more common due to restraints in nature and competition among species.

The results in this thesis tell us that diversification rate is, in fact, decreasing in most phylogenies and that it is because of a decay in the speciation rate. The number of species is still increasing but with a much slower rate. These results tell us that a possible limit to diversity could exist.

The thesis is structured as follows: First we will give a background to the process used in this thesis to illustrate the evolutionary process of our set of bird phylogenies. We will then continue by deriving the likelihood needed to estimate the speciation and extinction rates. Thirdly we will explain the different diversification scenarios that will be used to test the hypothesis that there is a limit to diversification. We have models where diversification rate declines as alternative hypothesis (model 3-6) against the null hypothesis where diversification rate remains constant through time (model 1 and 2). Analysis will first be done assuming we have all extant (now living) species represented in the phylogenetic tree and then assuming we only have a subsample of all extant species.

2 Background

Mathematical models are necessary to study the diversification rate. In this thesis birth-death models will be used, which are widely used in phylogenetics and the study of diversification. First, in section 2.1 a brief explanation of the birth and death process will be given. Then in section 2.2 we will go deeper into what is called the reconstructed process which is the process we will use to illustrate the phylogenetic tree. In section 2.3 we will for time dependant rates derive the likelihood function of the speciation times given the total age of the tree.

We will use speciation and birth, extinction and death, tree and phylogeny, as synonyms throughout the text.

2.1 Birth-death process

The birth-death process is a stochastic process having states 0,1,2... which represent the number of currently living species or individuals. When the process is in state n, it can in one transition either go to state $n-1$ (a death) or to state $n+1$ (a birth). A birth happens at an exponential rate $n\lambda$ and a death happens at an exponential rate $n\mu$. The waiting time until the next transition is exponentially distributed with rate $n(\lambda+\mu)$. These types of processes are well known in statistics and are called "continuous time Markov chains".

In a constant birth-death process each species can either speciate (give birth) with a constant rate λ or go extinct (die out) with a constant rate μ. Here the diversification rate will be speciation minus extinction $(\lambda-\mu)$.

The pure birth process (Yule, 1924) is a special case of the linear birth-death process with the extinction rate being 0. Hence, each species gives birth to a new one with a constant rate λ and extinction does not exist. With only one lineage the waiting time until the next birth is exponentially distributed with rate λ and with n lineages the rate is $n\lambda$.

2.2 The reconstructed process

The definition of the birth-death process presented in section 2.1 is the most widely used in phylogenetics but considers only constant rates over time. However, in this thesis the scenario where the rates vary over time will additionally be used, i.e. the birth and death rates are functions over time, $\lambda(t)$ and $\mu(t)$. Time varying rates are needed to study if the diversification rate is decreasing, which could indicate that there is a limit to diversity.

The data are trees from the species which have survived until present day. The construction of such trees, so-called reconstructed trees, is done by simulating a birth-death process and pruning all extinct lineages from what is called the complete tree (the third tree in Figure 2). Therefore we will start deriving the probabilities of the ordinary

birth-death process where extinct lineages are represented. Then we will modify the ordinary birth-death process by conditioning on that the process survives to present time and then remove every extinct lineage. This will generate what is called the "reconstructed" process and we will follow the derivations as done in Nee, May, & Harvey (1994). We will derive the distribution of the number of lineages in the reconstructed process and the likelihood function $L(\theta)$ for the reconstructed process given the total age of the tree.

We will assume that we have one single lineage at the beginning of the process $(t=0)$. The general birth-death process is a process that can either die or stay alive until some time t between 0 and present time T. From the first process we can form three others (Figure 2). The second process will survive until some arbitrary time t and may or may not go extinct before present time T. From the second process we will construct a third process using the information that the process survives to present time. The fourth process is the reconstructed process. This process is the same as the third except that we remove all the lineages that died before time T.
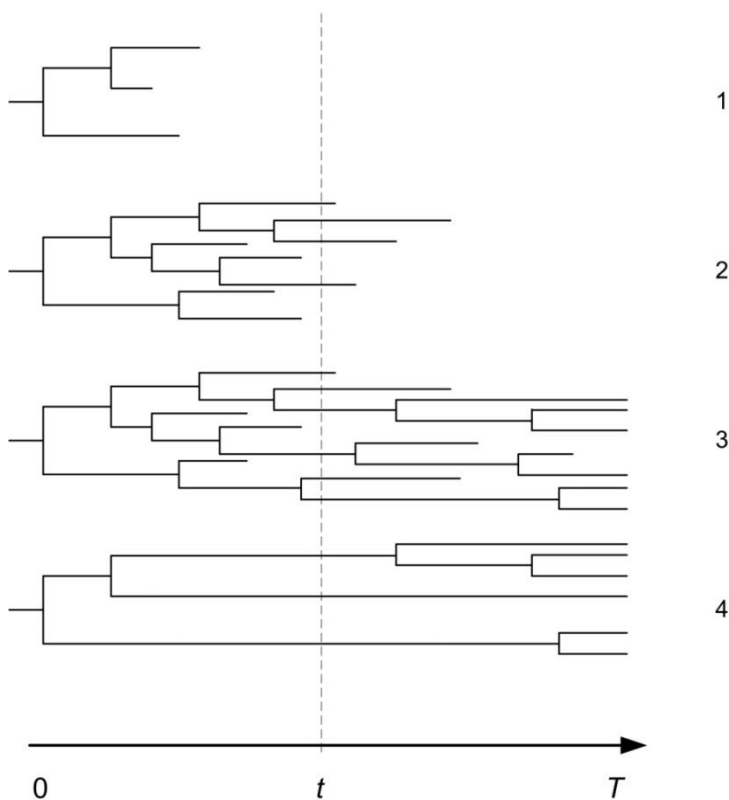

Figure 2. The first tree is generated from an ordinary birth-death process. The second tree is a birth-death process which survives to time t. Tree number 3 can be seen as a complete phylogenetic tree and survives to present time and tree number 4 is the reconstructed phylogeny of tree number 3 where every extinction event has been removed.

Let $P(n(t)=n)$ denote the probability of the number of lineages at time t, where $n(t)$ is the random variable which stands for the number of lineages at time t. Let the superscript of $P(n(t)=n)$ denote which process it belongs to. We start with process one which is an ordinary birth and death process. In a short amount of time dt three things can happen, a birth event, no event or a death event;

$$\begin{cases} n(t+dt)=n(t)+1 & \text{with probability } \lambda(t)n(t)dt \\ n(t+dt)=n(t) & \text{with probability } 1-(\lambda(t)+\mu(t))n(t)dt \\ n(t+dt)=n(t)-1 & \text{with probability } \mu(t)n(t)dt \end{cases} \tag{1}$$

At the starting time there is only have one lineage $n(0)=1$, the probability of having one lineage at time 0 is therefore $P(n(0)=1)=1$ and the probability of having $n \neq 1$ lineages at the beginning is $P(n(0)=n)=0$. Using these probabilities and the generating function Kendall (1948) obtain the probability function for the first process $P^1(n(t)=n)$, which can either die or stay alive until some arbitrary time t,

$$\begin{cases} P^1(n(t)=0)=1-P(0,t) \\ P^1(n(t)=n)=P(0,t)(1-u_t)u_t^{n-1} \end{cases} \tag{2}$$

Where $P(t,T)$ is the probability that a single lineage at time t is still alive at a later time T, given in Kendall (1948),

$$P(t,T)=\cfrac{1}{1+\displaystyle\int_t^T \exp(\rho(t,\tau))\mu(\tau)d\tau}, \tag{3}$$

where

$$\rho(t,\tau)=\int_t^\tau (\mu(s)-\lambda(s))ds.$$

And $1-u_t$ is the probability of a single offspring after a time t,

$$1-u_t = P(0,t)\exp(\rho(0,t)). \tag{4}$$

Both (3) and (4) will be used to form the likelihood function of the speciation times in the reconstructed process. Next, to be able to get to $P^2(n(t)=n)$ we only need to realize that the probability of having n species at a specific time in the second process is the same as $P^1(n(t)=n)$ with the exception that we know that it survives to time t, thus the term $P(0,t)$ in (2) is not needed,

$$P^2(n(t)=n)=(1-u_t)u_t^{n-1}. \tag{5}$$

5

In the third process at least one lineage survives to present time T. That is, it is the second process conditioned on that it survives until the present time T. Nevertheless, the probability densities of the third process are not necessary for the reconstructed process. The fourth process is the third except that we remove every extinction event. We obtain the fourth process if we do as in equation (1) and (2) but consider a birth process instead (no extinction exist). Each lineage gives birth to a new one with probability $\lambda(t)P(t,T)$. We know the process survives to at least time t but we want it to survive to the present time as well, and therefore we include the extra $P(t,T)$ which is the probability for a single lineage to survive to present time. After a small time interval either a birth event can occur or no event;

$$\begin{cases} n(t+dt) = n(t)+1 \text{ with probability } n(t)\lambda(t)P(t,T)dt \\ n(t+dt) = n(t) \quad \text{with probability } 1-n(t)\lambda(t)P(t,T)dt \end{cases} \quad (6)$$

Using these equations we get a probability function much the same the second equation in (2) except that $P(0,t)$ is not needed, because the process survives to time t. Instead of u_t we will have $u_t\,P(0,T)/P(0,t)$, where $P(0,T)/P(0,t)$ is the probability that a process that is alive at time t is still alive at T. We now have a geometric distribution with parameter $1 - u_t\,P(0,T)/P(0,t)$;

$$P^4\left(n(t)=n\right) = (1-u_t\frac{P(0,T)}{P(0,t)})(u_t\frac{P(0,T)}{P(0,t)})^{n-1}\ . \quad (7)$$

Equation (7) is the probability density function for the number of species alive at time t, given that at least one lineage is alive at time T. We could use (7) to calculate for example the expected number of species at any time point. If we would like to use a likelihood conditioned on the number of species instead of doing as here, where the likelihood is conditioned on the total age of the tree, probability function (7) is essential.

2.3 Likelihood function of speciation times
The data in molecular phylogenetics is the set of every i:th, $i=2...N$, speciation event $\{t_2,t_3,...,t_N\}$ (see figure 3). t_2 will in this situation be the time where the most recent common ancestor (MRCA) gave birth and defines the start of the clade, the origin.
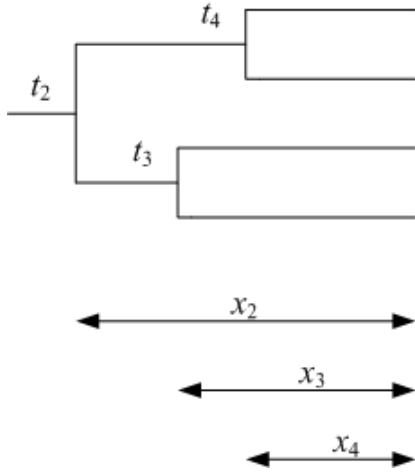
Figure 3 A phylogenetic tree. The t_i represent the time for the speciation events and the x_i represent the length of time from the nodes to present time.

Let $x_n = T - t_n$, x_n is the length of time between the present and the birth of the $n:$th lineage. The two outer lineages survives from time t_2 to time T with length x_2. Because we define t_2 as the origin, then $x_2 = T$. If we have n lineages in a tree there will be a probability of birth proportional to $n\lambda(t_{n+1})P(t_{n+1}, T)$, see (6), because of this the birth events in the tree contributes to the likelihood with

$$\lambda(t_2)P(t_2, T) 2\lambda(t_3)P(t_3, T)...(N-1)\lambda(t_N)P(t_N, T) = (N-1)! \prod_{i=2}^{N} \lambda(t_i)P(t_i, T). \qquad (8)$$

Equation (8) differs a little from the what they use in Nee, May & Harvey (1994). They do not include the first term $\lambda(t_2)P(t_2, T)$, the probability of the birth of the second lineage, in their likelihood. They believe that to be able to watch the phylogenetic tree the first speciation event have to occur and therefore $\lambda(t_2)P(t_2, T)$ does not give any contribution to the likelihood. We disagree with Nee, May & Harvey (1994), we do believe that this term contributes to the likelihood and therefore we choose to keep it in our model.

The other contribution to the likelihood comes from the amount of time the process does not give birth, i.e. the probability that the lineages observed at t_i has not split into two until present time. As said before in (4) the probability for a single lineage not to give birth in a given time x_i is $1 - u_{x_i}$. In the likelihood this will be

$$(1 - u_{x_2})^2 \prod_{i=3}^{N} (1 - u_{x_i}).$$

$(1 - u_{x_2})^2$ comes from the two outer lineages which arose from the first speciation event. The whole likelihood function of the speciation times for a phylogeny with N lineages

given the total age of the tree, with different parameters $\mathbf{\theta} = (\theta_1, ..., \theta_n)$ depending on how the rates $\lambda(t)$ and $\mu(t)$ are chosen, will therefore be

$$L(\mathbf{\theta}) = (N-1)! \left(\prod_{i=2}^{N} \lambda(t_i) P(t_i, T) \right) (1 - u_{x_2})^2 \prod_{i=3}^{N} (1 - u_{x_i}) =$$

$$(N-1)!(1 - u_{x_2}) \left(\prod_{i=2}^{N} \lambda(t_i) P(t_i, T) \right) \left(\prod_{i=2}^{N} 1 - u_{x_i} \right) =$$

$$(N-1)!(1 - u_{x_2}) \left(\prod_{i=2}^{N} \lambda(t_i) P(t_i, T)(1 - u_{x_i}) \right) =$$

$$(N-1)! P(t_2, T) \exp\left(\rho(t_2, T) \right) \left(\prod_{i=2}^{N} \lambda(t_i) P(t_i, T)^2 \exp\left(\rho(t_i, T) \right) \right). \tag{9}$$

3 Method

To be able to say if the diversification rate has decreased over time we will need to use different diversification models. In this section it will decide which models to use. We will derive the likelihood functions and then estimate birth and death rates for each model for a set of bird phylogenies from Phillimore and Price (2008). Section 3.1 contains the models where the assumption has been made that all the extant species of a clade are represented. In section 3.2 we deal with the problem of only having a random subsample of the extant species.

3.1 Diversification models with complete sampling

To answer our main question we will test several hypothesis models against our null models, the constant birth and death model and the constant birth model (model 1 and 2), see Rabosky (2006), who also used this approach. Different scenarios can lead to a declining diversification rate, which are discussed in the following. The birth rate is constant but the death rate increases with time, the death rate is constant but birth rate decreases with time or the birth rate decreases and death rate increases. We will construct the models given in table 1:

Model	Number of parameters	Model Properties	Parameters		Equations for Rates
Model 1	2	Constant birth and death rates	λ μ	birth rate death rate	$\lambda(t) = \lambda$ $\mu(t) = \mu$
Model 2	1	Constant Birth rate, no extinction (Yule)	λ	birth rate	$\lambda(t) = \lambda$
Model 3	2	Decreasing birth rate, no extinction	λ_0 α	initial birth rate variation in birth rate	$\lambda(t) = \lambda_0 \exp(-\alpha t)$
Model 4	4	Decreasing birth rate, increasing death rate	λ_0 α μ_T β	initial birth rate variation in birth rate death rate at present variation in death rate	$\lambda(t) = \lambda_0 \exp(-\alpha t)$ $\mu(t) = \mu_T (1 - \exp(-\beta t))$
Model 5	3	Constant birth rate, increasing death rate	λ μ_T β	birth rate death rate at present variation in death rate	$\lambda(t) = \lambda$ $\mu(t) = \mu_T (1 - \exp(-\beta t))$
Model 6	3	Decreasing birth rate, constant death rate	λ_0 α μ	initial birth rate variation in birth rate death rate	$\lambda(t) = \lambda_0 \exp(-\alpha t)$ $\mu(t) = \mu$

Table 1. The different diversification models with description of their properties and parameters.

As seen in table 1 we have chosen exponential increasing and decreasing rate models. These models have been presented in different experiments before (Rabosky D. L., 2008; Morlon, Potts, & Plotkin, 2010). All models represent expanding diversity but with different degrees. The speciation rate is believed to be greater than the extinction rate at all time[1]. Model 1, the constant birth and death process, will be reduced to model 2 (the Yule process) if the extinction rate is zero. In model 3 extinction is absent but the speciation rate is declining, thus giving declining diversification rate. If the parameter α, which denotes the magnitude of the rate decline, is equal to zero we will have a constant birth process, model 2. Model 6 is a simple extension of model 3 where a constant extinction rate is introduced. If $\alpha=0$ model 6 is reduced to model 1 with constant rates. Model 5 has asymptotically increasing extinction rate and constant speciation rate. The parameter β controls how fast the extinction rate increases towards its asymptotic value. If β is very large the extinction rate will be constant and the model will be reduced to the constant birth and death process. Model 4 has decreasing speciation rate and increasing extinction rate and can be reduced to any of the above given models. The different models are illustrated in figure 4.

---

[1] If $\int_0^t \lambda(t) - \mu(t)ds < 0$ for large t, then $P(n(t) = 0) = 1 - P(0,t) = 1$.
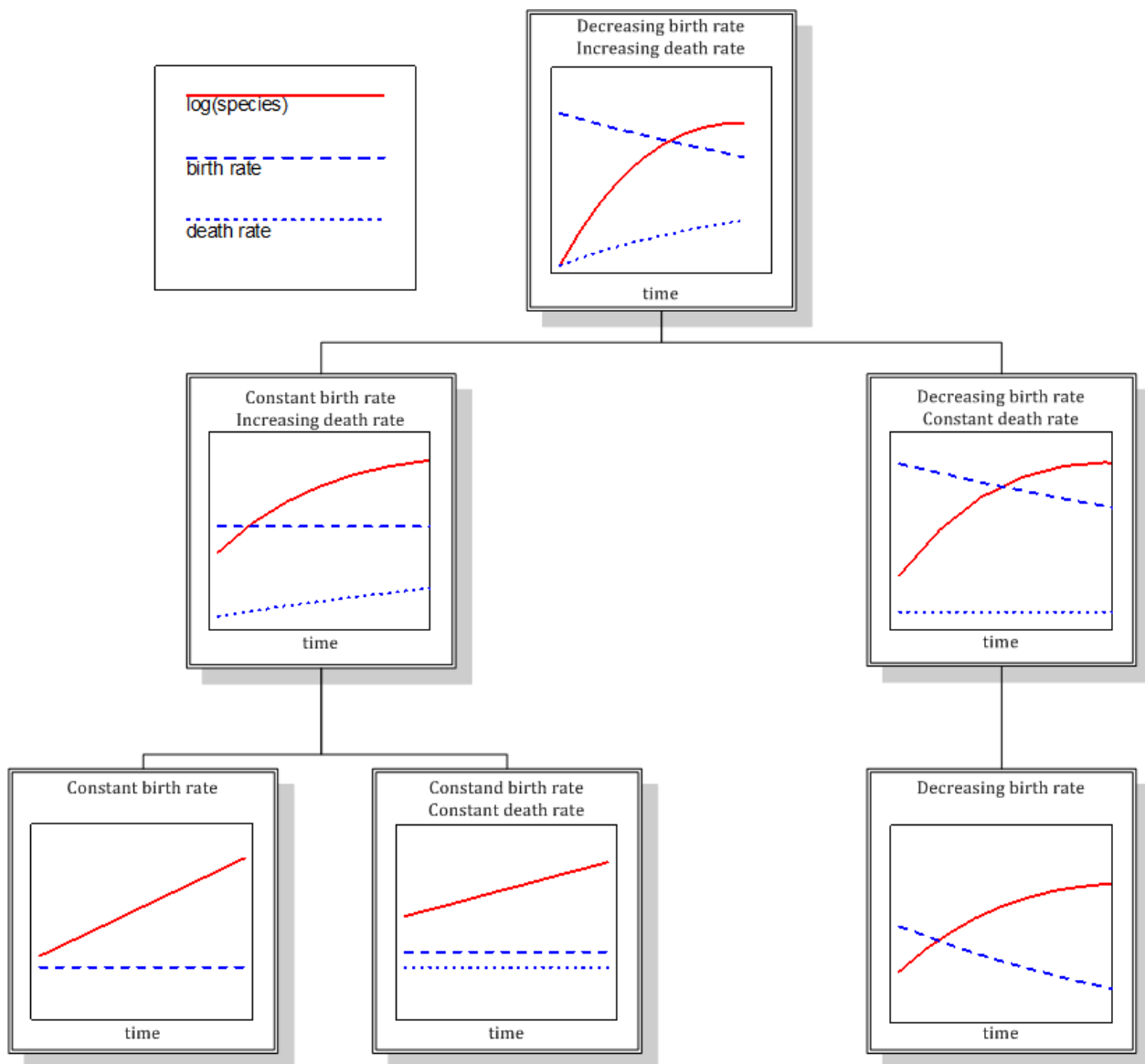
Figure 4. An illustration of the different diversification models explained in table 1. The red line symbolizes the natural logarithm of species under the different scenarios.

3.2 Diversification models with incomplete sampling

Real data consist often of incomplete samples (Nee, May, & Harvey, 1994; Stadler, 2009), i.e. the models are not based on all the living species in a clade but of a random sample of the living species. Therefore models accounting for incomplete taxon sampling will be included. The assumption of random sampling is not always met, but is easier to work with. A discussion is held in the end of section 5.2.

Incomplete taxon sampling, where every species has the same probability p of being sampled, can be modelled by a mass extinction at the time just before the species were observed (Nee, May & Harvey 1994). Let us assume that all species are included in the tree until very close to the present time T when a mass extinction occurs where each species has a probability p of surviving. We then get a reconstructed process where the

number of species used in our analysis is a subsample of all the extant species. To achieve this we use the Dirac delta function $\delta(x)$ and change the death rate into

$$\mu(t,T) = \mu(t) - \delta(t-T)\ln(p).$$

Here $\mu(t)$ is the death rate as before and $\delta(t-T)$ is the Dirac delta function with the property

$$\int_0^s \delta(t-T)dt = 0, \quad \text{if } s < T,$$

$$\int_0^s \delta(t-T)dt = 1, \quad \text{if } s > T$$

The Dirac delta function is zero everywhere except at $t = T$. Note, for technical implementation in R we need the property: for $\varepsilon > 0$

$$\int_{T-\varepsilon}^{T+\varepsilon} f(t)\delta(t-T)dt = f(T).$$

This will be applied on the same type of models as in table 1. In 20 of the 32 phylogenies there are species missing in the clade and are therefore suitable for this kind of analysis (see table of data in section 4.1).

3.3 Inference and model comparison

To make the maximum likelihood estimate for each parameter we first found the analytic solution for $\rho(t,\tau)$ for each model, see Appendix A4. For the time varying rate models we additionally needed to use numerical integration to find the solution to $P(t_i,T)$. We used integrate as implemented in R (Piessens, De Doncker-Kapenga, Überhuber, & Kahaner, 1983). This then enabled us to express the appropriate likelihood functions for every model. We used the Nelder-Mead method in the optimization function optim in R to do the estimations (Nelder & Mead, 1965).

To compare the fit of each model on a specific phylogeny we used the modified Akaike information criterion (Burnham & Anderson, 2004) which takes into account a finite sample size,

$$\text{AIC}_c = -2\log L(\hat{\theta}) + 2k + \frac{2k(k+1)}{n-k-1}.$$

Here $\log L(\hat{\theta})$ is the maximized log likelihood of the speciation times, k is the number of parameters in the model and n is the number of observations (the number of speciation events). The model with the lowest AIC$_c$ fit the data best. The first penalty, $2k$, is needed to reduce the overfitting problem in the likelihood when a model has many

parameters. The quotient is a adjustment to a finite sample and gives further penalty to models with many parameters and small data sets.

We also used the Akaike weights which may be interpreted as the probability that a given model is the best among a set of models. A model's Akaike weight (Burnham & Anderson, 2004), w_i, among a set with R models is calculated as follows

$$w_i = \frac{\exp(-\Delta_i / 2)}{\sum\limits_{r=1}^{R} \exp(-\Delta_r / 2)}.$$

Here Δ_i is the difference between the AIC$_c$ score of model i and the lowest AIC$_c$ score among all models. We compared the varying rate model with lowest AIC$_c$ score with the constant rate model with lowest AIC$_c$ score to determine with which degree a varying rate model fit the data compared to our null models with constant rates. The relative probability of the best model with varying rates compared to the best model with constant rates is

$$\frac{w_v}{w_v + w_c}.$$

The weight for the varying rate model is denoted w_v and the constant rate model weight is denoted w_c.

4 Analysis

4.1 The data

The data set constitutes of 32 different bird phylogenies from Phillimore and Price (2008). The phylogenies were generated from sequences of mitochondrial protein coding genes, using a relaxed clock Bayesian method implemented in BEAST (Phillimore & Price, 2008). In 20 of these phylogenies there are a known number of species missing, which have not yet been sequenced. The phylogenies, their size and clade age is given in Table 2.

Phylogeny	Number of species sampled	Number of species missing	Clade age (million years)
1. Aegotheles	8	1	10.221
2. Amazona	28	3	6.717
3. Anas	45	6	8.345
4. Anthus	37	9	12.653
5. Caciques and oropendolas	17	2	7.862
6. Dendroica, Parula, Seiurus, Vermivora	40	5	9.086
7. Grackles and allies	36	4	8.417
8. Hemispingus	12	2	15.688
9. Myiarchus	19	3	9.593
10. Phylloscopus and Seicercus	59	11	12.330
11. Puffinus	24	3	7.843
12. Ramphastos	8	3	8.114
13. Sterna	34	10	21.656
14. Storks	16	3	11.205
15. Tangara	42	5	10.104
16. Trogons	29	10	24.875
17. Turdus and allies	60	10	14.290
18. Wrens	50	24	12.098
19. Tringa	12	1	15.203
20. Meliphaga	12	1	15.645
21. Alectoris	7	0	7.950
22. Catharus	12	0	8.674
23. Cinclodes	13	0	5.190
24. Cranes	15	0	10.033
25. Crax	14	0	4.628
26. Penguins	18	0	10.011
27. Ficedula	27	0	15.772
28. Geositta	11	0	16.140
29. Albatross	14	0	9.080
30. Myioborus	12	0	4.707
31. Toxostoma	10	0	9.579
32. Acanthiza	13	0	9.903

Table 2. Summary table of the 32 bird phylogenies.

4.2 Result

The main result we have obtained is that the diversification rate declines over time and this is because of a decrease in the speciation rate. This diversification rate still leads to an expanding diversity but with a decreasing rate the closer it gets to present time. All phylogenies are best explained by a model with no extinction. In 21 of the 32 phylogenies (65.6%) model 3, the model with declining speciation rate and no extinction, fit the data best. In the remaining 11 phylogenies (34.4%) model 2, the constant birth model, has the lowest AIC_c score. Table 3 below gives the AIC_c score for each model and phylogeny. For extended results including the estimates, see appendix B.

Phylogeny	Model 1 cb-cd	Model 2 cb	Model 3 db	Model 4 db-id	Model 5 cb-id	Model 6 db-cd
1. Aegotheles						
p=1	31.387	27.187	27.846	48.846	38.387	34.846
p=0.8889	31.211	**27.011**	27.977	48.977	38.211	34.977
2. Amazona						
p=1	-10.097	-12.437	**-15.495**	-10.301	-7.554	-13.077
p=0.9032	-10.388	-12.728	-15.199	-10.051	-7.845	-12.796
3. Anas						
p=1	-77.948	**-79.067**	-76.869	-77.001	-77.719	-79.022
p=0.8824	-77.274	-77.805	-75.608	-76.425	-77.350	-78.416
4. Anthus						
p=1	21.153	18.907	**-4.868**	0.0592	23.539	-2.481
p=0.8043	19.342	17.096	-3.178	1.748	21.728	-0.792
5.Caciques and oropendolas						
p=1	28.113	25.476	**15.765**	22.363	31.190	18.740
p=0.8947	27.667	25.009	15.965	22.555	30.744	18.934
6. Dendroica, Parula, Seiurus, Vermivora						
p=1	-15.616	-17.841	**-29.229**	-24.566	-13.264	-27.057
p=0.8889	-16.317	-18.542	-28.681	-24.031	-13.964	-26.501
7. Grackles and allies						
p=1	-1.096	-3.349	**-23.584**	-18.840	1.304	-21.387
p=0.9	-1.976	-4.230	-23.206	-18.472	0.423	-21.019
8. Hemispingus						
p=1	41.460	38.404	**38.208**	47.188	45.388	41.994
p=0.8571	41.362	38.307	38.467	47.422	45.291	42.229
9. Myiarchus						
p=1	11.967	**9.417**	11.914	18.191	14.881	14.828
p=0.8636	12.164	9.614	12.163	18.440	15.079	15.077
10. Phylloscopus and Seicercus						
p=1	-31.618	-33.765	**-64.146**	-59.609	-29.392	-61.920
p=0.8429	-33.502	-35.649	-62.524	-57.988	-31.276	-60.298
11. Puffinus						
p=1	-2.333	-4.743	-3.273	2.350	0.3299	-0.6095
p=0.8889	-2.430	**-4.840**	-3.058	2.564	0.233	-0.395
12. Ramphastos						
p=1	27.818	**23.618**	25.454	46.387	34.818	32.416
p=0.7272	27.947	23.747	26.179	47.094	34.947	33.137
13. Sterna						
p=1	25.220	**22.949**	25.136	30.165	27.647	27.564
p=0.7727	27.0476	24.777	27.0477	32.070	29.469	29.475
14. Storks						
p=1	37.145	34.452	**30.110**	36.714	40.327	32.948
p=0.8421	36.851	34.158	30.549	37.087	40.032	33.330
15. Tangara						
p=1	-5.963	-8.176	**-29.343**	-24.540	-3.630	-27.002
p=0.8571	-7.262	-9.475	-28.367	-23.572	-4.929	-26.035
16. Trogons						
p=1	64.112	61.786	**56.538**	61.797	66.632	59.058
p=0.7436	64.034	61.708	58.829	64.089	66.554	61.349
17. Turdus and allies						
p=1	-54.032	-56.176	**-69.311**	-64.785	-51.810	-67.089
p=0.8571	-55.508	-57.653	-68.502	-63.976	-53.286	-66.280
18. Wrens						
p=1	4.113	1.937	**-40.581**	-35.932	6.386	-38.308
p=0.6757	1.453	-0.722	-33.456	-28.808	3.726	-31.183
19. Tringa						
p=1	43.636	40.581	**34.847**	43.013	47.565	38.775
p=0.9231	43.365	40.309	34.954	44.121	47.293	38.882
20. Meliphaga						
p=1	40.759	37.704	**36.306**	45.472	44.688	40.234
p=0.9231	40.526	37.471	36.386	45.553	44.455	40.315

Phylogeny						
21. *Alectoris* p=1	27.932	**22.932**	25.011	65.011	37.932	35.012
22. *Catharus* p=1	32.331	29.276	**23.021**	32.104	36.260	26.889
23. Cinclodes p=1	11.086	**8.153**	10.521	18.843	14.753	14.186
24. Cranes p=1	23.915	**21.157**	22.844	30.197	27.224	26.153
25. *Crax* p=1	14.019	11.182	**4.982**	12.782	17.485	8.448
26. Penguins p=1	23.249	**20.659**	22.253	28.345	26.238	25.009
27. *Ficedula* p=1	42.261	39.906	**32.789**	38.064	44.830	35.266
28. *Geositta* p=1	40.897	37.683	**36.491**	46.777	45.183	40.777
29. Albatross p=1	18.144	**15.311**	18.148	25.917	21.583	21.610
30. *Myioborus* p=1	14.744	11.689	**11.274**	20.440	18.673	15.202
31. *Toxostoma* p=1	32.704	29.276	**25.915**	37.915	37.504	30.715
32. *Acanthiza* p=1	34.329	31.396	**25.187**	33.568	37.996	28.854

Table 3. AIC$_c$ score for every phylogeny. Cb stands for constant birth rate, db for decreasing birth rate, cd for constant death rate and id for increasing death rate. The value highlighted is the model with lowest AIC$_c$ for each phylogeny.

The 21 phylogenies best explained by model 3, decreasing speciation rate and no extinction, is given in Table 4. The decrease in the speciation rate varies from 5.92 folds lower than the initial speciation rate (*Amazona* phylogeny) to 45,8 folds lower than the initial speciation rate (Wrens phylogeny). The mean and median of all the 21 phylogenies is 20 folds lower the initial rate. In 18 of the 21 phylogenies the present speciation rate is lower than 0.1 and in 14 phylogenies lower than 0.05.

Phylogeny	Initial speciation rate	Speciation rate today
2. *Amazona*	1.012	0.1708
4. *Anthus*	1.209	0.0397
5. Caciques and oropendolas	1.140	0.0374
6. *Dendroica, Parula, Seiurus, Vermivora*	0.9822	0.0919
7. Grackles and allies	1.309	0.0522
8. *Hemispingus*	0.3520	0.0442
10. *Phylloscopus* and *Seicercus*	1.129	0.0494
14. Storks	0.5905	0.0441
15. *Tangara*	1.1826	0.0565
16. Trogons	0.2784	0.0395
17. *Turdus* and allies	0.9635	0.1037
18. Wrens	1.210	0.0264
19. *Tringa*	0.600	0.0204
20. *Meliphaga*	0.4496	0.0450
22. *Catharus*	1.040	0.0258
25. *Crax*	1.988	0.0825
27. *Ficedula*	0.5333	0.0484
28. *Geositta*	0.4277	0.0389
30. *Myioborus*	1.225	0.1524
31. *Toxostoma*	0.8247	0.0314
32. *Acanthiza*	0.9155	0.0338

Table 4. The speciation rate when the process began and at present time in the 21 phylogenies best explained by model 3, decreasing speciation rate and no extinction.

All species are not represented in 20 of the phylogenies. This enables comparison between models taking random sampling into account and models with the assumption that all species are represented. In most of the phylogenies three of the six models taking into account random sampling get a better fit. The models that are improved are model 1, the constant birth-death model, model 2, the pure birth model and model 5, the model with constant birth rate and increasing death rate. This is, however, not the case in 20% of the phylogenies (the *Anas*, *Myiarchus, Ramphastos* and the *Sterna* phylogeny). Taking random sampling into account does not improve the model fit of the three models with decreasing birth rates, model 3, 4 and 6. This was not expected. We know that there is a subsample of species in these 20 phylogenies, and therefore we expected that the models assuming we have a subsample of species would get a better fit for these phylogenies.

We found that model 3, the model with decreasing birth rate and no extinction, assuming complete sampling explain the data best in most phylogenies. Conclusion drawn from models using the assumption that all species are represented could be false if the sample used for the model is incomplete, an impression of decline in the speciation rate could be the result of this procedure (Nee, Holmes, May, & Harvey, 1994; Rabosky D. L., 2008). When evaluating only the models which assumes random sampling we obtain the same main result of declining speciation rate as before. There is only two differences when only studying random sample models; model 6, decreasing speciation and constant extinction, gives a better fit instead of model 2 in the *Anas* phylogeny and model 2 get the lowest AIC_c score instead of model 3 in the *Hemispingus* phylogeny.

The results obtained when computing the Akaike weights for every phylogeny, even further strengthen our initial conclusion that there is a declining diversification rate. In 14 phylogenies the time varying model had more than 90% probability of being the true one compared to the best constant rate model, i.e. there is under 10% probability that the best model for these phylogenies is a constant rate model. In 12 of these phylogenies the varying rate model had more than 95% probability of being the best model explaining the phylogenies, see Fig. 5. The highest relative probability among the 11 phylogenies best explained by a constant rate model is just over 80%. For the 11 phylogenies best fit by the constant rate models the mean of the relative probability of being the true model is 70% compared to 88% for the 21 phylogenies best fit by varying rates models.
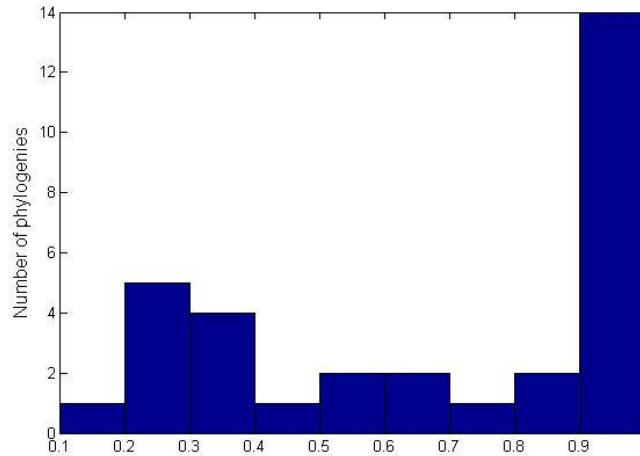
Figure 5. Relative probability of the best varying rate model compared to the best constant rate model for each phylogeny. The Akaike weights are used to compute the relative probability. The best varying rate model is in all cases except in the *Anas* phylogeny model 3. In the *Anas* phylogeny model 6 is the best varying rate model. Model 2, the Yule model, is in all phylogenies the best constant rate model to fit the data.

Studying the results in the appendix B we see that for model 1, the constant birth and death model, there are only 2 phylogenies where the extinction rate is not estimated to zero, both assuming complete sampling and random sampling. The same tendency is detected for model 5, the model with constant birth rate and increasing death rate, as for model 1. In model 4, decreasing birth rate and increasing death rate, there is 13 (40.6%) phylogenies where the extinction rate is not estimated to zero assuming complete sampling and 10 (50%) phylogenies where the extinction rate is not estimated to zero assuming random sampling. In both cases, with or without random sampling, all phylogenies with extinction rates estimated to zero where the same except for the *Sterna* phylogeny. The *Sterna* phylogeny had under complete sampling an extinction rate estimated to zero and under random sampling an estimate of 0.0088. The same results were obtained for model 6, the model with decreasing birth rate and constant death rate, when it comes to extinction rate, as for model 4.

We can also look at the natural logarithm of the number of species in different phylogenies. In Fig. 6 we have an example of four phylogenies; *Sterna*, *Puffinus*, Grackles and Trogons. The *Sterna* phylogeny is best explained by model 2, the constant birth process, assuming complete sampling and next best by model 2 assuming we have a random sample. The *Puffinus* phylogeny is also best explained by model 2, but the best model is the one assuming we have a random sample and next best assuming complete sampling. The Trogons and the Grackles phylogeny is best explained by model 3, the decreasing birth process, assuming complete sampling and next best by the corresponding model taking samples into account. We have plotted the logarithm of the expected number of species from these different models. We can see a difference of the curves representing the actual number of species in the trees best explained by a constant rate model (*Sterna* and *Puffinus*) compared to the trees best explained by the varying rate model (Trogons and Grackles).

The two best models explaining the Grackles phylogeny looks the same, the AIC$_c$ score in these two models are very similar, -23.584 and -23.21.
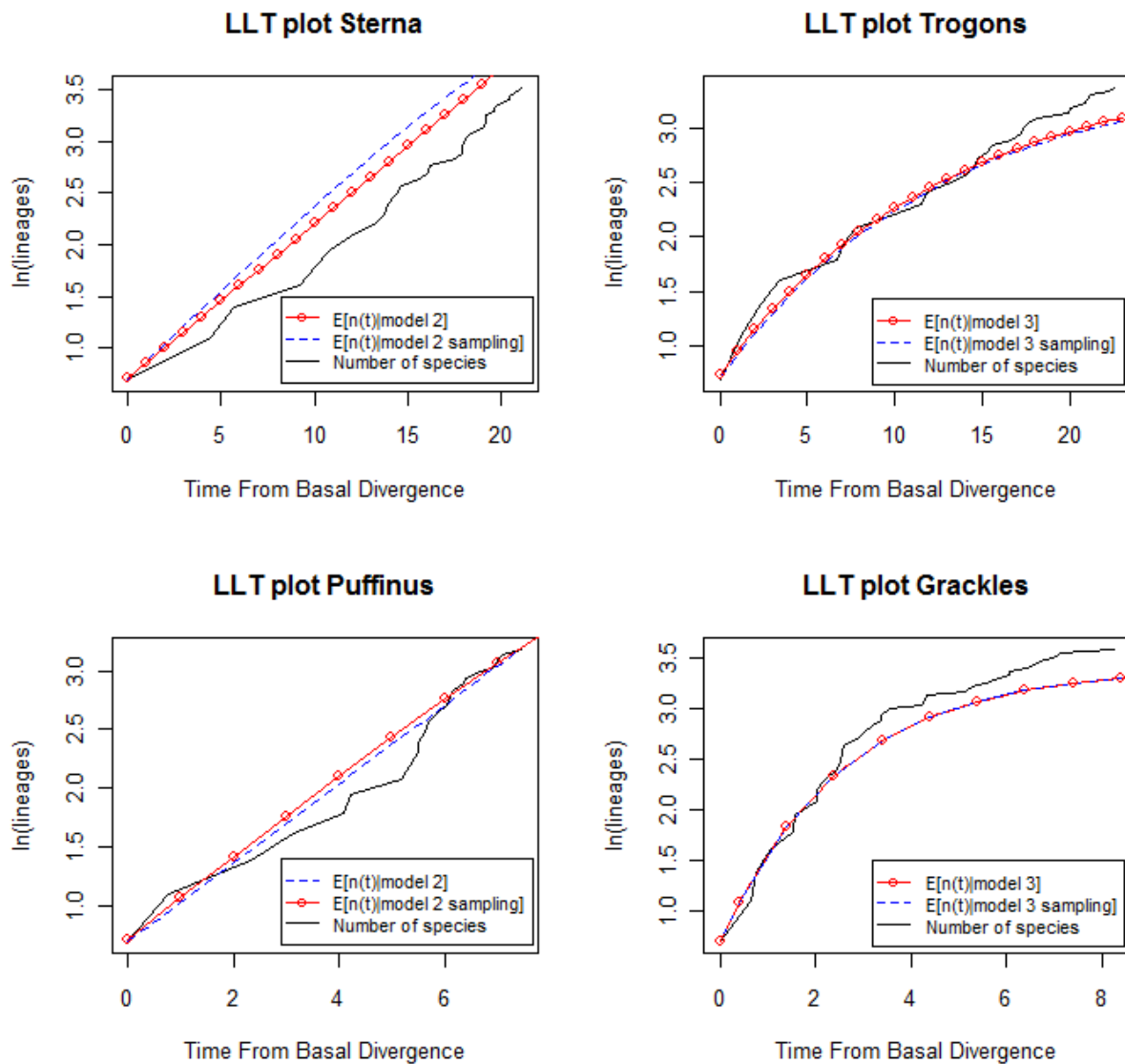
Figure 6. These lineages-through-time plots are of the natural logarithm of the actual number represented by the line, the expected number of species under the best model by the line with circles and the next best model by the dashed line. The Grackles and Trogons phylogeny is best explained by the pure decreasing birth rate model with complete sampling. The *Sterna* phylogeny is best explained by model 2, the constant birth process, assuming we have complete sampling and next best by the same model but assuming random sampling. The *Puffinus* phylogeny is also best explained by model 2 but with random sampling and next best by model 2 and complete sampling.

5 Discussion

Our results of declining diversification rates are in line with previous studies (Morlon, Potts, & Plotkin, 2010; Rabosky D. L., 2008). In Morlon, Potts, & Plotkin (2010) they use, among others, the same phylogenies as we do but use a different approach than the birth and death process. Their result is that most phylogenies are best explained by models with declining diversification rate. In Rabosky D. L. (2008) they use three phylogenies with a declining diversification rate and test whether it is changes in the speciation rate or extinction rate that is the cause. They use phylogenies with a sample fraction >0.93 and use models assuming complete sampling. They find that declining speciation rate and no extinction is the best-fit model in all three phylogenies. The declining speciation rate could be explained by adaptive radiation. In the past, species may have found relatively unoccupied niches and had the possibility to evolve in several different ways. Species then accumulated and adapted to the environment available to them, resulting in a slowdown in the speciation rate due to less resources and more competition (Gavrilets & Vose, 2005).

5.1 Extinction rate

The assumption used in this study that the speciation rate always is higher than the extinction rate may be an inaccurate representation of nature. When changing this assumption the results for the constant birth-death model are the same as before. This is, however, not very interesting because if the extinction rate would be higher than the speciation rate in the constant rate model the phylogenetic tree would, with high probability, die out before present time (Kendall, 1948). There are living species in the phylogenies and if the constant birth and death model is the underlying model it is likely that the speciation rate is higher than the extinction rate. However, it is relevant to change the assumption for the time varying rate models. A low extinction rate early in history could have resulted in the phylogenetic tree's survival, but later in time the extinction rate becomes higher than the speciation rate. When changing the assumption in model 5, constant birth rate and increasing death rate, there was only one case where the result differed from before. This was the *Anas* phylogeny and somewhere in its later history the extinction rate exceeds the speciation rate, it also leads to an improvement of the models fit.

When finding the parameters giving the maximum likelihood when removing the constraint that the extinction rate has to be larger than or equal to zero, we usually obtain a negative extinction rate (see Fig. 7). The *Anas* phylogeny is an exception and most of the times we get plots as the right one in Fig. 7, when plotting the extinction rate against the loglikelihood in the constant birth and death model.
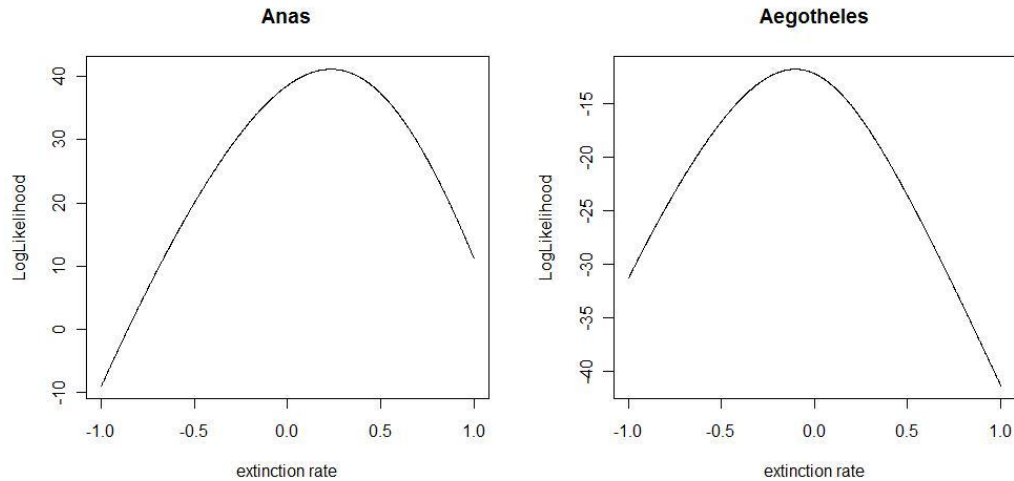
Figure 7. we fixed the value of the speciation rate giving the maximum log likelihood and made plots of the extinction rate.

5.2 Random sampling

Among the 20 phylogenies where species are missing, the two constant rate models had 4 phylogenies where the model assuming to have a random sample did not get a better fit than the corresponding model assuming that we have all species represented in the phylogeny. We also get the result that none of the models with random sampling and decreasing birth rate are better than the corresponding models assuming complete sampling. We have chosen to use a somewhat different equation for random sampling compared to Nee, May & Harvey (1994). They use a Taylor expansion of the component $\ln(p)$ into $p-1$ to obtain eq. 34 but we chose to keep $\ln(p)$ because it is an unnecessary approximation. When performing the analysis with this approximation one will get the result that all constant rate models with random sampling fit the data better. When using the approximation for model 3, decreasing birth rate and no extinction, we get that 60% of the phylogenies is best explained by the model under complete sampling. This leads to the conclusion that models with decreasing birth rate, of the exponential form in this thesis, and models with random sampling do not fit well together. Something must be changed if we want to continue exploring random sampling and varying rates together, especially if we want to use an exponential decreasing birth rate. To be able to establish if other decreasing models, and not only the exponentially decreasing models used in this thesis, has a problem with random sampling, we tried with a simple decreasing linear model with no extinction rate ($\lambda(t) = -\alpha t + \lambda_0$). We tried this on the first 16 phylogenies and they all gave the same result; the models with random sampling did not give a better fit. This is an important obstacle to overcome, because we do in fact know that there is a subsample of species in some of our data and we should include this in our analysis. This is important especially if we want to say with more confidence that the diversification rate has declined through time and is doing so because of a decreasing speciation rate in the incomplete phylogenies. There is a risk that the results obtained here is only the product of us not being able to model random sampling. We argued in section 4.2 that there can be an impression of a decline in the speciation rate

when having an incomplete sample but using models assuming complete sampling. To ascertain that the results obtained here is not a product of this mechanism we verified that the results would be the same if we only considered random sampling models. If declining speciation rate models under the assumption of random sampling does not work well together, and there is some relationship between them, then we cannot put full trust in these results. When studying only the phylogenies with a high sampling fraction $(0.9 < p)$, and using models assuming complete sampling, we get that 9 out of 14 (64.3%) phylogenies is best explained by model 3, the decreasing birth model. This leads us to the conclusion that our results before have some accuracy.

The sampling fraction could be another possible error source. To do these analyses we need to know how many species are missing to get the correct value of the sampling fraction p. If we missed some species that actually belong to the phylogeny, it will lead to overestimation of the sampling fraction, and our results would be incorrect.

One reason for the problem with the models under random sampling might be the way the mass extinction event is modelled. The lineages have been removed randomly in our analysis but in reality this may not always be true. To obtain the data in molecular phylogenetics we must have access to parts of species DNA. Even though sequencing of DNA is becoming easier and new species get their DNA sequenced continuously, it is still a matter of cost. This leads to that the choosing of which species to sequence could follow some pattern and not be random.

5.3 Starting the process at t_1 or t_2?

We do not know how long the first lineage lived before it split into two species. To be able to make our estimations, we set the first known speciation event as time 0. This is also done in Rabosky (2008) and is implemented in the LASER package in R (although with a different likelihood function) (Rabosky D. L., 2006:2). We wondered whether this would influence the outcome of decreasing birth rates not fitting well with random sampling. Therefore we also made the unknown time, from the birth of the very first species until the first splitting event, into a parameter. We tried to estimate it along with the other parameters. Unfortunately this did not work; we got the same estimations for the birth rate, death rate and the same likelihood as before for different values of this new parameter. Also, the problem remained that the models with decreasing birth rates with the assumption of incomplete sampling did not fit the data better than the models assuming complete sampling.

5.4 Studied models

Six different diversification scenarios have been tested for how well they explain different bird phylogenies. Two of the models are common null hypothesises in the study of cladogenesis. The other four were chosen because they are quite simple exponential models with few parameters but with opportunity to represent many different scenarios, including the two first ones. They have also been used in other studies and therefore give us the opportunity to compare our results. One could of

course use different time varying rate models than the ones used in this thesis. Even though the two models with no extinction best fit the bird phylogenies here, it is still important to use models with extinction present because we know that species have gone extinct throughout history.

6. Summary and future work

6.1 Summary and conclusions

We have studied how the diversification rate has changed through time in 32 bird phylogenies. This is done by using a modified birth-death process called the reconstructed process, describing the process of a phylogeny where the extinct species has been removed. Six different diversification scenarios, two with constant diversification rate and four with decreasing diversification rate, have been tested on the different phylogenies. The constant rate models are the constant birth-death model and the pure birth model (the Yule process). The declining diversification rate models are generated with decreasing speciation rate extinction, increasing extinction rate or both at the same time. We have also compared how well the models perform under the assumption that all extant species is represented in the phylogeny, named complete sampling compared to incomplete sampling. The result in this thesis indicates that the diversification rate is decreasing and is doing so because of a declining speciation rate. The two constant rate models and one of the declining diversification models perform better under the assumption of incomplete sampling. The other three declining diversification rate models, which do never performs better under incomplete sampling, have one thing in common; decreasing speciation rate. This is something which should be explored further.

6.2 Future work

Here, we have used an expression for the likelihood conditioned on the total age of the phylogenetic tree. There are two things we could change and continue working with. First, the total age contains some uncertainty regarding the time prior to the first speciation event and, second, we do know how many species are alive today and we can condition on that knowledge (Gernhard, 2008). We have written about it in appendix C. This is only for constant rates but if we could do it for time varying rates as well it would perhaps be a better way of tackling questions concerning diversification patterns.

Another thing one can try is to not let the speciation rate and extinction rate vary with time but rather with the clade size. If there only exist ten species in a clade the competition should not be so tough for space and resources and the extinction rate should be lower than if the clade contained a hundred species. This is called density dependant diversification and is discussed in Phillimore & Price (2008) and analysis is done in Seposki Jr (1978) and Rabosky & Lovette (2008). This would lead to a process with the properties

$$\begin{cases} n(t+dt) = n(t)+1 & \text{with probability } \lambda(n(t))n(t)dt \\ n(t+dt) = n(t) & \text{with probability } 1-\big(\lambda(n(t))+\mu(n(t))\big)n(t)dt \\ n(t+dt) = n(t)-1 & \text{with probability } \mu(n(t))n(t)dt \end{cases}$$
.

A third aspect to reconsider, which may be different in nature compared to these models and worth exploring, is that at a given time point each lineage has the same probability of going extinct and the same probability to speciate. Is this really likely? When competition hardens between species, is it not possible that one of the competitive species has the upper hand for a long time before the other goes extinct? If this is true then the species might not have the same probability of going extinct at a given time point.

Bibliography

Burnham, K. P., & Anderson, D. R. (2004, November). Multimodel Inference: Understanding AIC and BIC in Model Selection. *Sociological Methods Research* , pp. 261-304.

Gavrilets, S., & Vose, A. (2005, December 13). Dynamic patterns of adaptive radiation. *PNAS* , pp. 18040-18045.

Gernhard, T. (2008, August 21). The conditioned reconstructed process. *Journal of Theoretical Biology* , pp. 769-778.

Harvey, P. H., May, R. M., & Nee, S. (1994, June). Phylogenies Without Fossils. *Evolution* , pp. 523-529.

Jackman, T., Larson, A., De Queiroz, K., & Losos, B. J. (1999). Phylogenetic Relationships and Tempo of Early Diversification in Anoli Lizards. *Systematic Biology vol 48(2)* , 254-285.

Kendall, D. G. (1948). On the Generalized "Birth-and-Death" Process. *The Annals of Mathematical Statistics* , 1-15.

Kubo, T., & Iwasa, Y. (1995, August). Inferring the Rates of Brancing and Extinction from Molecular Phylogenies. *Evolution* , pp. 694-704.

Lieberman, B. S. (2001, August). A test of whether rates of speciation were unusually high during the Cambrian radiation. *Proceedings of the Royal Society vol.268* , pp. 1707-1714.

Morlon, H., Potts, M. D., & Plotkin, J. B. (2010, September). Inferring the Dynamics of Diversification: A Coalescent Approach. *PLoS Biology* .

Nee, S., Holmes, E. C., May, R. M., & H, H. P. (1994). Extinction rates can be estimated from molecular phylogenies. *Philosophical Transactions of the Royal Society: Biological Sciences vo.344l* , pp. 77-82.

Nee, S., May, R. M., & Harvey, P. H. (1994). The reconstruced evolutionary process. *Philosophical Transactions of the Royal Society of London* , 305-311.

Nelder, J. A., & Mead, R. (1965). A simplex algorithm for function minimization. *Computer Journal 7* , 308-313.

Phillimore, A. B., & Price, T. D. (2008, March). Density-Dependent Cladogenesis in Birds. *PLoS Biology* .

Piessens, R., De Doncker-Kapenga, E., Überhuber, C., & Kahaner, D. (1983). QUADPACK, A subroutine package for automatic integration. *Springer-Verlag* .

Rabosky, D. L. (2008, August). Explosive evolutionary radiations: decreasing speciation or increasing extinction through time? *Evolution* , pp. 1866-1875.

Rabosky, D. L. (2006:2). LASER: a maximum likelihood toolkit for detecting temporal shifts in diversification rates. *Evolutionary Bioinformatics Online* , 257-260.

Rabosky, D. L. (2006, June). Likelihood Methods for Detecting Temporal Shifts in Diversification Rates. *Evulotion* , pp. 1152-1164.

Rabosky, D. L., & Lovette, I. J. (2008, October). Density-dependent diversification in North American wood warbles. *Proceedings of the Royal Society Biological Sciences* , pp. 2363-2371 .

Seposki Jr, J. J. (1978). A Kinetic Model of Phanerozoic Taxonomic Diversity I. Analysis of Marine Orders. *Paleobiology Vol.4, No 3* , pp. 223-251.

Seposki, J. J. (1998). Rates of speciation in the fossil record. *Philosophical Transactions of the Royal Society vol. 353* , 315-326.

Stadler, T. (2009). On incomplete sampling under birth–death models and connections to the sampling-based coalescent. *Journal of Theoretical Biology* , 58-66.

Yang, Z., & Rannala, B. (1997). Bayesian phylogenetic inference using DNA sequences: A Markov chain Monte Carlo method. *Molecular Biology and Evolution 14 (7)* , pp. 717-724.

Yule, G. (1924). A Mathematical Theory of Evolution, Based on the Conclusions of Dr. J. C. Willis, F.R.S. *Philosophical Transactions of the Royal Society of London* , 21-87.

A   Equations

We will give a summary of all the most important equations as well as some extensions simplifying implementation in, for example, R.

A.1 Meanings

$\lambda(t)$ – the speciation rate

$\mu(t)$ – the extinction rate

$n(t)$ – number of species at time t

$P\big(n(t)=n\big)$ – the probability that a process has n lineages at time t

$P(t,T)$ – the probability that a single lineage at time t is still alive at a later time T

$1-u_t$ – the probability of a single offspring after a time t

t_i – the time of the $i:th$ birth event

x_n – the length of time between the present and the birth of the $n:th$ lineage

$\delta(x)$ – the dirac delta function

A.2 Complete sampling

For general rates we have

$$P(t,T)=\frac{1}{1+\displaystyle\int_{t}^{T}\exp(\rho(t,\tau))\mu(\tau)d\tau},$$

Where

$$\rho(t,\tau)=\int_{t}^{\tau}(\mu(s)-\lambda(s))ds.$$

For constant rate we have

$$P(t,T)=\frac{\lambda-\mu}{\lambda-\mu\exp(-(\lambda-\mu)(T-t))},$$

And commonly done in the literature we denote ρ by r when using constant rates

$r(t)=-(\lambda-\mu)t$ which leads to the expression

$$u_t=\frac{\lambda\big(1-\exp(r(t))\big)}{\lambda-\mu\exp(r(t))}.$$

A.3 Incomplete sampling

We will in the case of incomplete sampling have a different extinction rate due to the introduced mass extinction close to present time T. This will change both the ρ and the probability for a lineage to survive to present time,

$$\mu(t,T) = \mu(t) - \ln(p)\delta(t-T),$$

$$\rho(t,\tau) = \int_t^\tau \mu(s) - \lambda(s) - \ln(p)\delta(s-T)ds = \begin{cases} \int_t^\tau \mu(s) - \lambda(s)ds & \text{if } \tau < T \\ \int_t^\tau \mu(s) - \lambda(s)ds - \ln(p) & \text{if } \tau > T \end{cases}.$$

To calculate $P(t,T)$ we will also need to do the following

$$\int_t^T \exp(\rho(t,\tau))\mu(\tau)d\tau = \int_t^T \exp\left(\int_t^\tau \mu(s) - \lambda(s)ds - \ln(p)\Theta(\tau-T) \right)\mu(\tau)d\tau$$

$$= \int_t^T \frac{\exp\left(\int_t^\tau \mu(s) - \lambda(s)ds \right)\mu(\tau)}{\exp(\ln(p)\Theta(\tau-T))}d\tau,$$

where $\Theta(\tau-T)$ is 0 if $\tau < T$ and 1 if $\tau > T$.
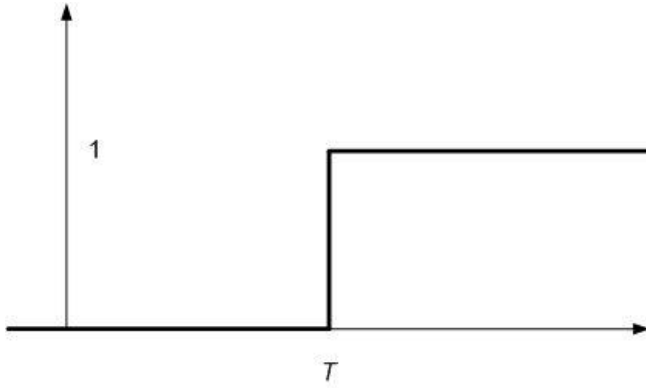
Figure 8. Plot of $\Theta(\tau-T)$

In the given integration interval the denominator will be equal to 1 and we have

$$\int_t^T \exp(\rho(t,\tau))\mu(\tau)d\tau = \int_t^T \exp\left(\int_t^\tau \mu(s) - \lambda(s)ds \right)\mu(\tau)d\tau.$$

With this and the property given in the main text,

$$\int_{T-\varepsilon}^{T+\varepsilon} f(t)\delta(t-T)dt = f(T),$$

we can simplify $P(t,T)$ with consideration to implementation in R.

$$P(t,T) = \cfrac{1}{1 + \int\limits_t^T \exp\big(\rho(t,\tau)\big)\big(\mu(\tau) - \ln(p)\delta(\tau - T)\big)d\tau}$$

$$= \cfrac{1}{1 + \int\limits_t^T \exp\big(\rho(t,\tau)\big)\mu(\tau)d\tau - \int\limits_t^T \ln(p)\exp\big(\rho(t,\tau)\big)\delta(\tau - T)d\tau}$$

$$= \cfrac{1}{1 + \int\limits_t^T \exp\left(\int\limits_t^\tau \mu(s) - \lambda(s)ds\right)\mu(\tau)d\tau - \ln(p)\exp\big(\rho(t,T)\big)}.$$

A.4 Likelihood functions for each model

The general likelihood function (9) is

$$(N-1)!\,P(t_2,T)\exp\big(\rho(t_2,T)\big)\left(\prod_{i=2}^N \lambda(t_i)P(t_i,T)^2\,\exp\big(\rho(t_i,T)\big)\right),$$

and we will in the following section give the corresponding $L(\boldsymbol{\theta})$, $P(t,T)$ and $\rho(t,T)$ to each model.

1. Constant birth rate and constant death rate.

$\lambda(t) = \lambda$ and $\mu(t) = \mu$

If we have complete sampling the equations will take the following form;

$$\rho(t,T) = \int\limits_t^T \mu - \lambda\,ds = (\mu - \lambda)(T - t)$$

$$P(t,T) = \frac{\lambda - \mu}{\lambda - \mu\exp\big(-(\lambda - \mu)(T - t)\big)}$$

If we have incomplete sampling we will instead have;

$$\rho(t,T) = \int\limits_t^T \mu - \lambda - \ln(p)\delta(s - T)ds = (\mu - \lambda)(T - t)\ - \ln(p)$$

$$P(t,T) = \cfrac{1}{1 + \mu\int\limits_t^T \exp\big((\mu - \lambda)(\tau - t)\big)d\tau - \exp\big(\rho(t,T)\big)\ln(p)}$$

$$= \cfrac{1}{1 + \cfrac{\mu}{\mu - \lambda} \left(\exp\left((\mu - \lambda)(T - t) \right) - 1 \right) - \cfrac{\ln(p)}{p} \exp\left((\mu - \lambda)(T - t) \right)}$$

$$= \frac{p(\lambda - \mu)}{p\lambda - \left(p\mu + (\lambda - \mu)\ln(p) \right) \exp\left((\mu - \lambda)(T - t) \right)}.$$

A Taylor expansion of the component $\ln(p)$ into $p - 1$ will generate the same equation as equation 34 in (Nee, May, & Harvey, 1994).

The likelihood be will be

$$L(\lambda,\mu) = (N-1)! \lambda^{N-1} P(t_2, T) \exp\left(\rho(t_2, T) \right) \prod_{i=2}^{N} P(t_i, T)^2 \exp\left(\rho(t_i, T) \right).$$

2. The constant birth model.

$\lambda(t) = \lambda$ and $\mu(t) = 0$

With complete sampling we have

$$\rho(t,T) = \int_t^T -\lambda \, ds = -\lambda(T - t).$$

The probability of surviving until present time is $P(t,T) = 1$ because no extinction exists.

Incomplete sampling yields,

$$\rho(t,T) = \int_t^T -\lambda - \ln(p)\delta(s - T) ds = -\lambda(T - t) - \ln(p)$$

$$P(t,T) = \frac{1}{1 - \ln(p)\exp(\rho(t,T))}.$$

The likelihood will for the constant birth model be

$$L(\lambda) = (N-1)! \lambda^{N-1} P(t_2, T) \exp(\rho(t_2, T)) \left(\prod_{i=2}^{N} P(t_i, T)^2 \exp\left(\rho(t_i, T) \right) \right).$$

In the case of complete sampling we have

$$\hat{\lambda} = \frac{N - 1}{x_2 + \displaystyle\sum_{i=2}^{N} x_i}$$

3. Decreasing birth rate with no extinction.

$\lambda(t) = \lambda_0 \exp(-\alpha t)$ and $\mu(t) = 0$

Assuming complete sampling gives

$$\rho(t,T) = \int_t^T -\lambda_0 \exp(-\alpha s) ds = \frac{\lambda_0}{\alpha}\left(\exp(-\alpha T) - \exp(-\alpha t)\right),$$

and the probability of surviving until present time is equal to 1 as in the previous model,

$$P(t,T) = \frac{1}{1 + \int_t^T \exp(\rho(t,\tau))\mu(\tau)d\tau} = 1.$$

Incomplete sampling yields,

$$\rho(t,T) = \int_t^T -\lambda_0 \exp(-\alpha s) - \ln(p)\delta(s-T) ds = \frac{\lambda_0}{\alpha}\left(\exp(-\alpha T) - \exp(-\alpha t)\right) - \ln(p),$$

$$P(t,T) = \frac{1}{1 - \ln(p)\int_t^T \exp(\rho(t,\tau))\delta(t-T)d\tau} = \frac{1}{1 - \ln(p)\exp(\rho(t,T))}.$$

The likelihood for the decreasing birth model will be

$$L(\lambda_0, \alpha) = (N-1)! \lambda_0^{N-1} P(t_2,T) \exp\left(\rho(t_2,T)\right)\left(\prod_{i=2}^N P(t_i,T)^2 \exp(-\alpha t_i)\exp\left(\rho(t_i,T)\right)\right).$$

4. Decreasing birth rate, increasing death rate.

$\lambda(t) = \lambda_0 \exp(-\alpha t)$ and $\mu(t) = \mu_T(1 - \exp(-\beta t))$

With complete sampling we have

$$\rho(t,T) = \int_t^T \mu_T(1 - \exp(-\beta s)) - \lambda_0 \exp(-\alpha s) ds =$$

$$= \mu_T(T-t) + \frac{\mu_T}{\beta}\left(\exp(-\beta T) - \exp(-\beta t)\right) + \frac{\lambda_0}{\alpha}\left(\exp(-\alpha T) - \exp(-\alpha t)\right).$$

The probability of surviving until present time is

$$P(t,T) = \frac{1}{1 + \mu_T \int_t^T \exp(\rho(t,\tau))(1 - \exp(-\beta \tau))d\tau}.$$

Incomplete sampling yields the following equations instead

$$\rho(t,T) = \mu_T(T-t) + \frac{\mu_T}{\beta}\big(\exp(-\beta T) - \exp(-\beta t)\big) + \frac{\lambda_0}{\alpha}\big(\exp(-\alpha T) - \exp(-\alpha t)\big) - \ln(p),$$

$$P(t,T) = \frac{1}{1 + \mu_T \int\limits_t^T \exp\left(\int\limits_t^\tau \mu(s) - \lambda(s)ds\right)\big(1 - \exp(-\beta s)\big)d\tau - \exp\big(\rho(t,T)\big)\ln(p)}.$$

The likelihood becomes

$$L(\lambda_0, \mu_T, \alpha, \beta) = (N-1)!\lambda_0^{N-1}P(t_2,T)\exp\big(\rho(t_2,T)\big)\left(\prod_{i=2}^N \exp(-\alpha t_i)P(t_i,T)^2 \exp\big(\rho(t_i,T)\big)\right).$$

5. Constant birth rate, increasing death rate.

$\lambda(t) = \lambda_0$ and $\mu(t) = \mu_T(1 - \exp(-\beta t))$

Under the assumption of complete sampling we will have these expressions;

$$\rho(t,T) = \int\limits_t^T \mu_T(1 - \exp(-\beta t)) - \lambda_0 ds = (\mu_T - \lambda_0)(T-t) + \frac{\mu_T}{\beta}\big(\exp(-\beta T) - \exp(-\beta t)\big),$$

$$P(t,T) = \frac{1}{1 + \mu_T \int\limits_t^T \exp(\rho(t,\tau))(1 - \exp(-\beta\tau))d\tau}.$$

Incomplete sampling gives the following instead;

$$\rho(t,T) = (\mu_T - \lambda_0)(T-t) + \frac{\mu_T}{\beta}\big(\exp(-\beta T) - \exp(-\beta t)\big) - \ln(p),$$

$$P(t,T) = \frac{1}{1 + \mu_T \int\limits_t^T \exp\left(\int\limits_t^\tau \mu(s) - \lambda(s)ds\right)\big(1 - \exp(-\beta s)\big)d\tau - \exp\big(\rho(t,T)\big)\ln(p)}.$$

The likelihood becomes

$$L(\lambda_0, \mu_T, \beta) = (N-1)!\lambda_0^{N-1}P(t_2,T)\exp\big(\rho(t_2,T)\big)\left(\prod_{i=2}^N P(t_i,T)^2 \exp\big(\rho(t_i,T)\big)\right).$$

6. Decreasing birth rate, constant death rate

$\lambda(t) = \lambda_0 \exp(-\alpha t)$ and $\mu(t) = \mu$

Under the assumption of complete sampling we will have these expressions;

$$\rho(t,T) = \int_t^T \mu - \lambda_0(\exp(-\alpha s))ds = \mu(T-t) + \frac{\lambda_0}{\alpha}\left(\exp(-\alpha T) - \exp(-\alpha t)\right),$$

$$P(t,T) = \frac{1}{1 + \mu\int_t^T \exp(\rho(t,\tau))d\tau}.$$

Incomplete sampling gives the following instead;

$$\rho(t,T) = \mu(T-t) + \frac{\lambda_0}{\alpha}\left(\exp(-\alpha T) - \exp(-\alpha t)\right),$$

$$P(t,T) = \frac{1}{1 + \mu\int_t^T \exp\left(\int_t^\tau \mu(s) - \lambda(s)ds\right)d\tau - \exp(\rho(t,T))\ln(p)}.$$

The likelihood will be

$$L(\lambda_0,\mu,\alpha) = (N-1)!\lambda_0^{N-1}P(t_2,T)\exp(\rho(t_2,T))\left(\prod_{i=2}^N \exp(-\alpha t_i)P(t_i,T)^2 \exp(\rho(t_i,T))\right).$$

B Table of results

Table of results complete taxon sampling

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
1. Aegotheles N=9 n=8	$\hat{\lambda}$=0.1409 $\hat{\mu} \approx 0$ LogL= -12.194	$\hat{\lambda}$=0.1409 LogL= -12.194	$\hat{\lambda}$=0.5656 $\hat{\alpha}$=0.2481 LogL= -10.423	$\hat{\lambda}$=0.5656 $\hat{\alpha}$=0.2481 $\hat{\mu} \approx 0$ $\hat{\beta}$=0.1089 LogL= -10.423	$\hat{\lambda}$=0.1409 $\hat{\mu} \approx 0$ $\hat{\beta}$=0.2129 LogL= -12.194	$\hat{\lambda}$=0.5656 $\hat{\alpha}$=0.2481 $\hat{\mu} \approx 0$ LogL= -10.423
2. Amazona N=31 n=28	$\hat{\lambda}$=0.3261 $\hat{\mu} \approx 0$ LogL= 7.299	$\hat{\lambda}$=0.3261 LogL= 7.299	$\hat{\lambda}$=1.012 $\hat{\alpha}$=0.2649 LogL= 9.997	$\hat{\lambda}$=1.297 $\hat{\alpha}$=0.2823 $\hat{\mu}$=0.1597 $\hat{\beta}$=0.3536 LogL= 10.06	$\hat{\lambda}$=0.3261 $\hat{\mu} \approx 0$ $\hat{\beta}$=3.5740 LogL= 7.299	$\hat{\lambda}$=1.448 $\hat{\alpha}$=0.2990 $\hat{\mu}$=0.1572 LogL=10.060
3. Anas N=51 n=45	$\hat{\lambda}$=0.5286 $\hat{\mu}$=0.2327 LogL= 41.120	$\hat{\lambda}$=0.3962 LogL= 40.581	$\hat{\lambda}$=0.3962 $\hat{\alpha} \approx 0$ LogL= 40.581	$\hat{\lambda}$=0.9772 $\hat{\alpha}$=0.0525 $\hat{\mu}$=0.8991 $\hat{\beta}$=0.1448 LogL= 43.019	$\hat{\lambda}$=0.6539335 $\hat{\mu}$=0.6539331 $\hat{\beta}$=0.1769 LogL= 42.159	$\hat{\lambda}$=1.434 $\hat{\alpha}$=0.1036 $\hat{\mu}$=0.6043 LogL= 42.811
4. Anthus N=46 n=37	$\hat{\lambda}$=0.1508 $\hat{\mu} \approx 0$ LogL= -8.394	$\hat{\lambda}$=0.1508 LogL= -8.394	$\hat{\lambda}$=1.209 $\hat{\alpha}$=0.2700 LogL= 4.616	$\hat{\lambda}$=1.209 $\hat{\alpha}$=0.2700 $\hat{\mu} \approx 0$ $\hat{\beta}$=3.811 LogL= 4.616	$\hat{\lambda}$=0.1508 $\hat{\mu} \approx 0$ $\hat{\beta}$=3.282 LogL= -8.394	$\hat{\lambda}$=1.209 $\hat{\alpha}$=0.2700 $\hat{\mu} \approx 0$ LogL= 4.616
5. Caciques and oropendolas N=19 n=17	$\hat{\lambda}$=0.1937 $\hat{\mu} \approx 0$ LogL= -11.595	$\hat{\lambda}$=0.1937 LogL= -11.595	$\hat{\lambda}$=1.140 $\hat{\alpha}$=0.4348 LogL= -5.421	$\hat{\lambda}$=1.267 $\hat{\alpha}$=0.4362 $\hat{\mu}$=0.0411 $\hat{\beta}$=0.9736 LogL= -5.363	$\hat{\lambda}$=0.1937 $\hat{\mu} \approx 0$ $\hat{\beta}$=1.918 LogL= -11.595	$\hat{\lambda}$=1.281 $\hat{\alpha}$=0.4387 $\hat{\mu}$=0.0407 LogL= -5.370
6. Dendroica, Parula, Seiurus, Vermivora N=45 n=40	$\hat{\lambda}$=0.2280 $\hat{\mu} \approx 0$ LogL= 9.975	$\hat{\lambda}$=0.2280 LogL= 9.975	$\hat{\lambda}$=0.9822 $\hat{\alpha}$=0.2607 LogL= 16.781	$\hat{\lambda}$=1.317 $\hat{\alpha}$=0.2771 $\hat{\mu}$=0.1029 $\hat{\beta}$=3.978 LogL= 16.871	$\hat{\lambda}$=0.2280 $\hat{\mu} \approx 0$ $\hat{\beta}$=14.246 LogL= 9.975	$\hat{\lambda}$=1.305 $\hat{\alpha}$=0.2769 $\hat{\mu}$=0.0988 LogL= 16.862
7. Grackles and allies N=40 n=36	$\hat{\lambda}$=0.2113 $\hat{\mu} \approx 0$ LogL= 2.735	$\hat{\lambda}$=0.2113 LogL= 2.735	$\hat{\lambda}$=1.309 $\hat{\alpha}$=0.3827 LogL= 13.980	$\hat{\lambda}$=1.504 $\hat{\alpha}$=0.3850 $\hat{\mu}$=0.0591 $\hat{\beta}$=0.6654 LogL= 14.087	$\hat{\lambda}$=0.2113 $\hat{\mu} \approx 0$ $\hat{\beta}$=1.870 LogL= 2.735	$\hat{\lambda}$=1.5324 $\hat{\alpha}$=0.3882 $\hat{\mu}$=0.0584 LogL= 14.081

Taxon	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
8. Hemispingus N=14 n=12	$\hat{\lambda}$=0.1080 $\hat{\mu}\approx0$ LogL= -17.980	$\hat{\lambda}$=0.1080 LogL= -17.980	$\hat{\lambda}$=0.3520 $\hat{\alpha}$=0.1316 LogL= -16.354	$\hat{\lambda}$=0.4240 $\hat{\alpha}$=0.1318 $\hat{\mu}$=0.0785 $\hat{\beta}$=0.0732 LogL= -16.261	$\hat{\lambda}$=0.1080 $\hat{\mu}\approx0$ $\hat{\beta}$=3.127 LogL= -17.980	$\hat{\lambda}$=0.4766 $\hat{\alpha}$=0.1408 $\hat{\mu}$=0.0524 LogL= -16.283
9. Myiarchus N=22 n=19	$\hat{\lambda}$=0.2949 $\hat{\mu}\approx0$ LogL= -3.583	$\hat{\lambda}$=0.2949 LogL= -3.583	$\hat{\lambda}$=0.3420 $\hat{\alpha}$=0.0218 LogL= -3.557	$\hat{\lambda}$=0.3420 $\hat{\alpha}$=0.0218 $\hat{\mu}\approx0$ $\hat{\beta}$=0.5615 LogL=-3.557	$\hat{\lambda}$=0.2949 $\hat{\mu}\approx0$ $\hat{\beta}$=3.053 LogL= -3.583	$\hat{\lambda}$=0.3420 $\hat{\alpha}$=0.0218 $\hat{\mu}\approx0$ LogL= -3.557
10. Phylloscopus and Seicercus N=70 n=59	$\hat{\lambda}$=0.1649 $\hat{\mu}\approx0$ LogL= 17.918	$\hat{\lambda}$=0.1649 LogL= 17.918	$\hat{\lambda}$=1.129 $\hat{\alpha}$=0.2537 LogL= 34.182	$\hat{\lambda}$=1.1289 $\hat{\alpha}$=0.2537 $\hat{\mu}\approx0$ $\hat{\beta}$=0.1151 LogL=34.182	$\hat{\lambda}$=0.1649 $\hat{\mu}\approx0$ $\hat{\beta}$=1.385 LogL= 17.918	$\hat{\lambda}$=1.1291 $\hat{\alpha}$=0.2537 $\hat{\mu}\approx0$ LogL= 34.182
11. Puffinus N=27 n=24	$\hat{\lambda}$=0.3352 $\hat{\mu}\approx0$ LogL=3.467	$\hat{\lambda}$=0.3352 LogL= 3.467	$\hat{\lambda}$=0.5694 $\hat{\alpha}$=0.0980 LogL= 3.936	$\hat{\lambda}$=0.5694 $\hat{\alpha}$=0.0980 $\hat{\mu}\approx0$ $\hat{\beta}$=1.100 LogL=3.936	$\hat{\lambda}$=0.3352 $\hat{\mu}\approx0$ $\hat{\beta}$=4.083 LogL= 3.467	$\hat{\lambda}$=0.5694 $\hat{\alpha}$=0.0980 $\hat{\mu}\approx0$ LogL= 3.936
12. Ramphastos N=11 n=8	$\hat{\lambda}$=0.1818 $\hat{\mu}\approx0$ LogL= -10.409	$\hat{\lambda}$=0.1818 LogL= -10.409	$\hat{\lambda}$=0.5731 $\hat{\alpha}$=0.2596 LogL= -9.227	$\hat{\lambda}$=0.7423 $\hat{\alpha}$=0.2769 $\hat{\mu}$=0.0785 $\hat{\beta}$=2.1396 LogL=-9.194	$\hat{\lambda}$=0.1818 $\hat{\mu}\approx0$ $\hat{\beta}$=2.000 LogL= -10.409	$\hat{\lambda}$=0.7491 $\hat{\alpha}$=0.2799 $\hat{\mu}$=0.0773 LogL= -9.208
13. Sterna N=44 n=34	$\hat{\lambda}$=0.1506 $\hat{\mu}\approx0$ LogL= -10.410	$\hat{\lambda}$=0.1506 LogL= -10.410	$\hat{\lambda}$=0.1756 $\hat{\alpha}$=0.0097 LogL= -10.368	$\hat{\lambda}$=0.1756 $\hat{\alpha}$=0.0097 $\hat{\mu}\approx0$ $\hat{\beta}$=1.100 LogL=-10.368	$\hat{\lambda}$=0.1506 $\hat{\mu}\approx0$ $\hat{\beta}$=2.354 LogL= -10.410	$\hat{\lambda}$=0.1757 $\hat{\alpha}$=0.0097 $\hat{\mu}\approx0$ LogL= -10.368
14. Storks N=19 n=16	$\hat{\lambda}$=0.1449 $\hat{\mu}\approx0$ LogL= -16.072	$\hat{\lambda}$=0.1449 LogL= -16.072	$\hat{\lambda}$=0.5905 $\hat{\alpha}$=0.2316 LogL= -12.555	$\hat{\lambda}$=0.7088 $\hat{\alpha}$=0.2334 $\hat{\mu}$=0.0520 $\hat{\beta}$=0.5066 LogL=-12.3572	$\hat{\lambda}$=0.1449 $\hat{\mu}\approx0$ $\hat{\beta}$=3.004 LogL= -16.072	$\hat{\lambda}$=0.7302 $\hat{\alpha}$=0.2380 $\hat{\mu}$=0.0507 LogL= -12.383
15. Tangara N=49 n=42	$\hat{\lambda}$=0.1909 $\hat{\mu}\approx0$ LogL= 5.139	$\hat{\lambda}$=0.1909 LogL= 5.139	$\hat{\lambda}$=1.1826 $\hat{\alpha}$=0.3010 LogL= 16.825	$\hat{\lambda}$=1.1826 $\hat{\alpha}$=0.3010 $\hat{\mu}\approx0$ $\hat{\beta}$=3.224 LogL=16.825	$\hat{\lambda}$=0.1909 $\hat{\mu}\approx0$ $\hat{\beta}$=1.810 LogL= 5.139	$\hat{\lambda}$=1.182 $\hat{\alpha}$=0.3010 $\hat{\mu}\approx0$ LogL= 16.825

16. Trogons N=39 n=29	$\hat{\lambda}$=0.0830 $\hat{\mu} \approx 0$ LogL= -29.816	$\hat{\lambda}$=0.0830 LogL= -29.816	$\hat{\lambda}$=0.2784 $\hat{\alpha}$=0.0785 LogL= -26.029	$\hat{\lambda}$=0.2784 $\hat{\alpha}$=0.0785 $\hat{\mu} \approx 0$ $\hat{\beta}$=0.0359 LogL=-26.029	$\hat{\lambda}$=0.0830 $\hat{\mu} \approx 0$ $\hat{\beta}$=3.219 LogL= -29.816	$\hat{\lambda}$=0.2784 $\hat{\alpha}$=0.0785 $\hat{\mu} \approx 0$ LogL= -26.029
17. *Turdus* and allies N=70 n=60	$\hat{\lambda}$=0.1951 $\hat{\mu} \approx 0$ LogL= 29.123	$\hat{\lambda}$=0.1951 LogL= 29.123	$\hat{\lambda}$=0.9635 $\hat{\alpha}$=0.1560 LogL= 36.763	$\hat{\lambda}$=0.9635 $\hat{\alpha}$=0.1560 $\hat{\mu} \approx 0$ $\hat{\beta}$=0.1125 LogL=36.763	$\hat{\lambda}$=0.1951 $\hat{\mu} \approx 0$ $\hat{\beta}$=1.5892 LogL= 29.123	$\hat{\lambda}$=0.9636 $\hat{\alpha}$=0.1560 $\hat{\mu} \approx 0$ LogL= 36.763
18. Wrens N=74 n=50	$\hat{\lambda}$=0.1424 $\hat{\mu} \approx 0$ LogL= 0.0738	$\hat{\lambda}$=0.1424 LogL= 0.0738	$\hat{\lambda}$=1.210 $\hat{\alpha}$=0.3162 LogL= 22.421	$\hat{\lambda}$=1.210 $\hat{\alpha}$=0.3162 $\hat{\mu} \approx 0$ $\hat{\beta}$=2.0106 LogL=22.421	$\hat{\lambda}$=0.1424 $\hat{\mu} \approx 0$ $\hat{\beta}$=4.840 LogL= 0.0738	$\hat{\lambda}$=1.210 $\hat{\alpha}$=0.3162 $\hat{\mu} \approx 0$ LogL= 22.421
19. *Tringa* N=13 n=12	$\hat{\lambda}$=0.0978 $\hat{\mu} \approx 0$ LogL= -19.068	$\hat{\lambda}$=0.0978 LogL= -19.068	$\hat{\lambda}$=0.6000 $\hat{\alpha}$=0.2223 LogL= -14.673	$\hat{\lambda}$=0.6000 $\hat{\alpha}$=0.2223 $\hat{\mu} \approx 0$ $\hat{\beta}$=0.8921 LogL= -14.673	$\hat{\lambda}$=0.0978 $\hat{\mu} \approx 0$ $\hat{\beta}$=2.437 LogL= -19.068	$\hat{\lambda}$=0.6001 $\hat{\alpha}$=0.2223 $\hat{\mu} \approx 0$ LogL= -14.673
20. *Meliphaga* N=13 n=12	$\hat{\lambda}$=0.1115 $\hat{\mu} \approx 0$ LogL= -17.630	$\hat{\lambda}$=0.1115 LogL= -17.630	$\hat{\lambda}$=0.4496 $\hat{\alpha}$=0.1471 LogL= -15.403	$\hat{\lambda}$=0.4494 $\hat{\alpha}$=0.1471 $\hat{\mu} \approx 0$ $\hat{\beta}$=1.334 LogL= -15.403	$\hat{\lambda}$=0.1115 $\hat{\mu} \approx 0$ $\hat{\beta}$=2.435 LogL= -17.630	$\hat{\lambda}$=0.4495 $\hat{\alpha}$=0.1472 $\hat{\mu} \approx 0$ LogL= -15.403
21. *Alectoris* N=7 n=7	$\hat{\lambda}$=0.1725 $\hat{\mu} \approx 0$ LogL= -9.966	$\hat{\lambda}$=0.1725 LogL= -9.966	$\hat{\lambda}$=0.6473 $\hat{\alpha}$=0.3104 LogL= -8.505	$\hat{\lambda}$=0.6473 $\hat{\alpha}$=0.3104 $\hat{\mu} \approx 0$ $\hat{\beta}$=01.216 LogL= -8.505	$\hat{\lambda}$=0.1724 $\hat{\mu} \approx 0$ $\hat{\beta}$=2.564 LogL= -9.966	$\hat{\lambda}$=0.6472 $\hat{\alpha}$=0.3104 $\hat{\mu} \approx 0$ LogL= -8.505
22. *Catharus* N=12 n=12	$\hat{\lambda}$=0.1635 $\hat{\mu} \approx 0$ LogL= -13.416	$\hat{\lambda}$=0.1635 LogL= -13.416	$\hat{\lambda}$=1.040 $\hat{\alpha}$=0.4262 LogL= -8.760	$\hat{\lambda}$=1.080 $\hat{\alpha}$=0.4192 $\hat{\mu}$=0.2236 $\hat{\beta}$=0.0157 LogL= -8.719	$\hat{\lambda}$=0.1635 $\hat{\mu} \approx 0$ $\hat{\beta}$=1.613 LogL= -13.416	$\hat{\lambda}$=1.134 $\hat{\alpha}$=0.4254 $\hat{\mu}$=0.0283 LogL= -8.730

23. **Cinclodes** N=13 n=13	$\hat{\lambda}$=0.4044 $\hat{\mu} \approx 0$ LogL= -2.876	$\hat{\lambda}$=0.4044 LogL= -2.876	$\hat{\lambda}$=0.6699 $\hat{\alpha}$=0.1527 LogL= -2.594	$\hat{\lambda}$=1.043 $\hat{\alpha}$=0.2145 $\hat{\mu}$=0.2362 $\hat{\beta}$=2.035 LogL= -2.564	$\hat{\lambda}$=0.4044 $\hat{\mu} \approx 0$ $\hat{\beta}$=3.340 LogL= -2.876	$\hat{\lambda}$=0.7514 $\hat{\alpha}$=0.1705 $\hat{\mu}$=0.0517 LogL= -2.593
24. **Cranes** N=15 n=15	$\hat{\lambda}$=0.2295 $\hat{\mu} \approx 0$ LogL= -9.412	$\hat{\lambda}$=0.2295 LogL= -9.412	$\hat{\lambda}$=0.4589 $\hat{\alpha}$=0.1024 LogL= -8.876	$\hat{\lambda}$=0.4590 $\hat{\alpha}$=0.1024 $\hat{\mu} \approx 0$ $\hat{\beta}$=0.835 LogL=-8.876	$\hat{\lambda}$=0.2295 $\hat{\mu} \approx 0$ $\hat{\beta}$=1.671 LogL= -9.412	$\hat{\lambda}$=0.4590 $\hat{\alpha}$=0.1024 $\hat{\mu} \approx 0$ LogL= -8.876
25. **Crax** N=14 n=14	$\hat{\lambda}$=0.3417 $\hat{\mu} \approx 0$ LogL= -4.410	$\hat{\lambda}$=0.3417 LogL= -4.410	$\hat{\lambda}$=1.988 $\hat{\alpha}$=0.6877 LogL= 0.1091	$\hat{\lambda}$=1.988 $\hat{\alpha}$=0.6877 $\hat{\mu} \approx 0$ $\hat{\beta}$=0.3284 LogL=0.1091	$\hat{\lambda}$=0.3417 $\hat{\mu} \approx 0$ $\hat{\beta}$=1.957 LogL= -4.410	$\hat{\lambda}$=1.988 $\hat{\alpha}$=0.6877 $\hat{\mu} \approx 0$ LogL= 0.1091
26. **Penguins** N=18 n=18	$\hat{\lambda}$=0.2205 $\hat{\mu} \approx 0$ LogL= -9.196	$\hat{\lambda}$=0.2205 LogL= -9.196	$\hat{\lambda}$=0.3789 $\hat{\alpha}$=0.0849 LogL= -8.748	$\hat{\lambda}$=0.7233 $\hat{\alpha}$=0.1287 $\hat{\mu} \approx 0.1996$ $\hat{\beta}$=0.7538 LogL=-8.506	$\hat{\lambda}$=0.2205 $\hat{\mu} \approx 0$ $\hat{\beta}$=3.461 LogL= -9.196	$\hat{\lambda}$=0.7571 $\hat{\alpha}$=0.1364 $\hat{\mu}$=0.1932 LogL= -8.581
27. **Ficedula** N=27 n=27	$\hat{\lambda}$=0.1247 $\hat{\mu} \approx 0$ LogL= -18.870	$\hat{\lambda}$=0.1247 LogL= -18.870	$\hat{\lambda}$=0.5333 $\hat{\alpha}$=0.1521 LogL= -14.133	$\hat{\lambda}$=0.7095 $\hat{\alpha}$=0.1616 $\hat{\mu}$=0.0546 $\hat{\beta}$=2.542 LogL= -14.079	$\hat{\lambda}$=0.1247 $\hat{\mu} \approx 0$ $\hat{\beta}$=4.132 LogL= -18.870	$\hat{\lambda}$=0.6980 $\hat{\alpha}$=0.1612 $\hat{\mu}$=0.0507 LogL= -14.088
28. **Geositta** N=11 n=11	$\hat{\lambda}$=0.1034 $\hat{\mu} \approx 0$ LogL= -17.591	$\hat{\lambda}$=0.1034 LogL= -17.591	$\hat{\lambda}$=0.4277 $\hat{\alpha}$=0.1485 LogL= -15.388	$\hat{\lambda}$=0.4277 $\hat{\alpha}$=0.1485 $\hat{\mu} \approx 0$ $\hat{\beta}$=0.3072 LogL= -15.388	$\hat{\lambda}$=0.1034 $\hat{\mu} \approx 0$ $\hat{\beta}$=0.9047 LogL= -17.591	$\hat{\lambda}$=0.4278 $\hat{\alpha}$=0.1485 $\hat{\mu} \approx 0$ LogL= -15.388
29. **Albatross** N=14 n=14	$\hat{\lambda}$=0.2992 $\hat{\mu}$=0.0172 LogL= -6.472	$\hat{\lambda}$=0.2915 LogL= -6.474	$\hat{\lambda}$=0.2915 $\hat{\alpha} \approx 0$ LogL= -6.474	$\hat{\lambda}$=0.3153 $\hat{\alpha} \approx 0$ $\hat{\mu}$=0.0535 $\hat{\beta}$=0.9298 LogL= -6.458	$\hat{\lambda}$=0.3153 $\hat{\mu}$=0.0534 $\hat{\beta}$=0.9297 LogL= -6.458	$\hat{\lambda}$=0.2992 $\hat{\alpha} \approx 0$ $\hat{\mu}$=0.0171 LogL= -6.472
30. **Myioborus** N=12 n=12	$\hat{\lambda}$=0.3637 $\hat{\mu} \approx 0$ LogL= -4.622	$\hat{\lambda}$=0.3637 LogL= -4.622	$\hat{\lambda}$=1.225 $\hat{\alpha}$=0.4428 LogL= -2.887	$\hat{\lambda}$=1.225 $\hat{\alpha}$=0.4428 $\hat{\mu} \approx 0$ $\hat{\beta}$=0.481 LogL= -2.887	$\hat{\lambda}$=0.3637 $\hat{\mu} \approx 0$ $\hat{\beta}$=0.1135 LogL= -4.622	$\hat{\lambda}$=1.225 $\hat{\alpha}$=0.4428 $\hat{\mu} \approx 0$ LogL=-2.887

31. Toxostoma N=10 n=10	$\hat{\lambda}$=0.1487 $\hat{\mu} \approx 0$ LogL= -13.352	$\hat{\lambda}$=0.1487 LogL= -13.352	$\hat{\lambda}$=0.8247 $\hat{\alpha}$=0.3411 LogL= -9.958	$\hat{\lambda}$=0.8247 $\hat{\alpha}$=0.3411 $\hat{\mu} \approx 0$ $\hat{\beta}$=0.0887 LogL= -9.958	$\hat{\lambda}$=0.1487 $\hat{\mu} \approx 0$ $\hat{\beta}$=0.6054 LogL= -13.352	$\hat{\lambda}$=0.8246 $\hat{\alpha}$=0.3411 $\hat{\mu} \approx 0$ LogL= -9.958
32. Acanthiza N=13 n=13	$\hat{\lambda}$=0.1535 $\hat{\mu} \approx 0$ LogL= -14.498	$\hat{\lambda}$=0.1535 LogL= -14.498	$\hat{\lambda}$=0.9155 $\hat{\alpha}$=0.3332 LogL= -9.027	$\hat{\lambda}$=0.9155 $\hat{\alpha}$=0.3332 $\hat{\mu} \approx 0$ $\hat{\beta}$=0.2279 LogL= -9.027	$\hat{\lambda}$=0.1535 $\hat{\mu} \approx 0$ $\hat{\beta}$=2.613 LogL= -14.498	$\hat{\lambda}$=0.9155 $\hat{\alpha}$=0.3332 $\hat{\mu} \approx 0$ LogL= -9.927

Table of results random taxon sampling

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
1. Aegotheles N=9 n=8	$\hat{\lambda}$=0.1537 $\hat{\mu} \approx 0$ LogL= -12.106	$\hat{\lambda}$=0.1537 LogL= -12.106	$\hat{\lambda}$=0.5684 $\hat{\alpha}$=0.2341 LogL= -10.488	$\hat{\lambda}$=0.5684 $\hat{\alpha}$=0.2341 $\hat{\mu} \approx 0$ $\hat{\beta}$=1.107 LogL= -10.488	$\hat{\lambda}$=0.1537 $\hat{\mu} \approx 0$ $\hat{\beta}$=0.7622 LogL= -12.106	$\hat{\lambda}$=0.5684 $\hat{\alpha}$=0.2341 $\hat{\mu} \approx 0$ LogL= -10.488
2. Amazona N=31 n=28	$\hat{\lambda}$=0.3468 $\hat{\mu} \approx 0$ LogL= 7.444	$\hat{\lambda}$=0.3468 LogL= 7.444	$\hat{\lambda}$=1.001 $\hat{\alpha}$=0.2483 LogL= 9.849	$\hat{\lambda}$=1.490 $\hat{\alpha}$=0.2850 $\hat{\mu}$=0.1845 $\hat{\beta}$=4.964 LogL= 9.935	$\hat{\lambda}$=0.3468 $\hat{\mu} \approx 0$ $\hat{\beta}$=0.7610 LogL= 7.444	$\hat{\lambda}$=1.460 $\hat{\alpha}$=0.2838 $\hat{\mu}$=0.1717 LogL=9.920
3. Anas N=51 n=45	$\hat{\lambda}$=0.6035 $\hat{\mu}$=0.3076 LogL=40.783	$\hat{\lambda}$=0.4197 LogL= 39.950	$\hat{\lambda}$=0.4197 $\hat{\alpha} \approx 0$ LogL= 39.950	$\hat{\lambda}$=1.029 $\hat{\alpha}$=0.0416 $\hat{\mu}$=1.029 $\hat{\beta}$=0.1470 LogL= 42.725	$\hat{\lambda}$=0.7458931 $\hat{\mu}$=0.7458928 $\hat{\beta}$=0.1994 LogL= 41.975	$\hat{\lambda}$=1.534 $\hat{\alpha}$=0.095 $\hat{\mu}$=0.6933 LogL= 42.508
4. Anthus N=46 n=37	$\hat{\lambda}$=0.1763 $\hat{\mu} \approx 0$ LogL= -7.489	$\hat{\lambda}$=0.1763 LogL= -7.489	$\hat{\lambda}$=1.165 $\hat{\alpha}$=0.2445 LogL= 3.771	$\hat{\lambda}$=1.165 $\hat{\alpha}$=0.2445 $\hat{\mu} \approx 0$ $\hat{\beta}$=0.2145 LogL= 3.771	$\hat{\lambda}$=0.1763 $\hat{\mu} \approx 0$ $\hat{\beta}$=2.659 LogL= -7.489	$\hat{\lambda}$=1.165 $\hat{\alpha}$=0.2445 $\hat{\mu} \approx 0$ LogL= 3.771
5. Caciques and oropendolas N=19 n=17	$\hat{\lambda}$=0.2091 $\hat{\mu} \approx 0$ LogL= -11.372	$\hat{\lambda}$=0.2091 LogL= -11.372	$\hat{\lambda}$=1.153 $\hat{\alpha}$=0.4197 LogL= -5.521	$\hat{\lambda}$=1.295 $\hat{\alpha}$=0.4210 $\hat{\mu}$=0.0473 $\hat{\beta}$=0.9883 LogL= -5.459	$\hat{\lambda}$=0.2091 $\hat{\mu} \approx 0$ $\hat{\beta}$=4.177 LogL= -11.372	$\hat{\lambda}$=1.311 $\hat{\alpha}$=0.4237 $\hat{\mu}$=0.0469 LogL= -5.467

6. Dendroica, Parula, Seiurus, Vermivora N=45 n=40	$\hat{\lambda}$=0.2454 $\hat{\mu} \approx 0$ LogL= 10.325	$\hat{\lambda}$=0.2454 LogL= 10.325	$\hat{\lambda}$=0.9720 $\hat{\alpha}$=0.2461 LogL= 16.507	$\hat{\lambda}$=1.313 $\hat{\alpha}$=0.2624 $\hat{\mu}$=0.1100 $\hat{\beta}$=3.891 LogL= 16.603	$\hat{\lambda}$=0.2454 $\hat{\mu} \approx 0$ $\hat{\beta}$=2.284 LogL= 10.325	$\hat{\lambda}$=1.301 $\hat{\alpha}$=0.2622 $\hat{\mu}$=0.1055 LogL= 16.593
7. Grackles and allies N=40 n=36	$\hat{\lambda}$=0.2266 $\hat{\mu} \approx 0$ LogL= 3.175	$\hat{\lambda}$=0.2266 LogL= 3.175	$\hat{\lambda}$=1.308 $\hat{\alpha}$=0.3685 LogL= 13.980	$\hat{\lambda}$=1.529 $\hat{\alpha}$=0.3713 $\hat{\mu}$=0.0672 β=0.7576 LogL= 13.903	$\hat{\lambda}$=0.2266 $\hat{\mu} \approx 0$ $\hat{\beta}$=0.3820 LogL= 2.735	$\hat{\lambda}$=1.556 $\hat{\alpha}$=0.3744 $\hat{\mu}$=0.0666 LogL= 13.790
8. Hemispingus N=14 n=12	$\hat{\lambda}$=0.1190 $\hat{\mu} \approx 0$ LogL= -17.931	$\hat{\lambda}$=0.1190 LogL= -17.931	$\hat{\lambda}$=0.3565 $\hat{\alpha}$=0.1222 LogL= -16.484	$\hat{\lambda}$=0.4348 $\hat{\alpha}$=0.1211 $\hat{\mu}$=0.1055 $\hat{\beta}$=0.0612 LogL= -16.377	$\hat{\lambda}$=0.1190 $\hat{\mu} \approx 0$ $\hat{\beta}$=2.100 LogL= -17.931	$\hat{\lambda}$=0.5016 $\hat{\alpha}$=0.1317 $\hat{\mu}$=0.635 LogL= -16.400
9. Myiarchus N=22 n=19	$\hat{\lambda}$=0.3221 $\hat{\mu} \approx 0$ LogL= -3.6821	$\hat{\lambda}$=0.3221 LogL= -3.6821	$\hat{\lambda}$=0.3290 $\hat{\alpha}$=0.0031 LogL= -3.6816	$\hat{\lambda}$=0.3291 $\hat{\alpha}$=0.0032 $\hat{\mu} \approx 0$ $\hat{\beta}$=0.8880 LogL=-3.6816	$\hat{\lambda}$=0.3221 $\hat{\mu} \approx 0$ $\hat{\beta}$=2.941 LogL= -3.6821	$\hat{\lambda}$=0.3291 $\hat{\alpha}$=0.0032 $\hat{\mu} \approx 0$ LogL= -3.6816
10. Phylloscopus and Seicercus N=70 n=59	$\hat{\lambda}$=0.1848 $\hat{\mu} \approx 0$ LogL= 18.860	$\hat{\lambda}$=0.1848 LogL= 18.860	$\hat{\lambda}$=1.103 $\hat{\alpha}$=0.2359 LogL= 33.371	$\hat{\lambda}$=1.103 $\hat{\alpha}$=0.2359 $\hat{\mu} \approx 0$ $\hat{\beta}$=0.2681 LogL=33.371	$\hat{\lambda}$=0.1848 $\hat{\mu} \approx 0$ $\hat{\beta}$=3.101 LogL= 18.860	$\hat{\lambda}$=1.103 $\hat{\alpha}$=0.2359 $\hat{\mu} \approx 0$ LogL= 33.371
11. Puffinus N=27 n=24	$\hat{\lambda}$=0.3605 $\hat{\mu} \approx 0$ LogL=3.515	$\hat{\lambda}$=0.3605 LogL= 3.515	$\hat{\lambda}$=0.5528 $\hat{\alpha}$=0.0791 LogL= 3.829	$\hat{\lambda}$=0.5528 $\hat{\alpha}$=0.0791 $\hat{\mu} \approx 0$ $\hat{\beta}$=0.2810 LogL=3.829	$\hat{\lambda}$=0.3605 $\hat{\mu} \approx 0$ $\hat{\beta}$=1.232 LogL= 3.515	$\hat{\lambda}$=0.5528 $\hat{\alpha}$=0.0791 $\hat{\mu} \approx 0$ LogL= 3.829
12. Ramphastos N=11 n=8	$\hat{\lambda}$=0.2269 $\hat{\mu} \approx 0$ LogL= -10.474	$\hat{\lambda}$=0.2269 LogL= -10.474	$\hat{\lambda}$=0.5937 $\hat{\alpha}$=0.2187 LogL= -9.5894	$\hat{\lambda}$=0.8360 $\hat{\alpha}$=0.2393 $\hat{\mu}$=0.1199 $\hat{\beta}$=2.223 LogL=-9.547	$\hat{\lambda}$=0.2269 $\hat{\mu} \approx 0$ $\hat{\beta}$=1.589 LogL= -10.474	$\hat{\lambda}$=0.8452 $\hat{\alpha}$=0.2434 $\hat{\mu}$=0.1173 LogL= -9.568
13. Sterna N=44 n=34	$\hat{\lambda}$=0.1751 $\hat{\mu}$=0.0011 LogL= -11.324	$\hat{\lambda}$=0.1745 LogL= -11.324	$\hat{\lambda}$=0.1745 $\hat{\alpha} \approx 0$ LogL= -11.324	$\hat{\lambda}$=0.1791 $\hat{\alpha} \approx 0$ $\hat{\mu}$=0.0088 $\hat{\beta}$=0.5106 LogL=-11.321	$\hat{\lambda}$=0.1790 $\hat{\mu}$=0.0087 $\hat{\beta}$=0.5094 LogL= -11.321	$\hat{\lambda}$=0.1751 $\hat{\alpha} \approx 0$ $\hat{\mu}$=0.0012 LogL= -11.324

14. Storks N=19 n=16	$\hat{\lambda}$=0.1622 $\hat{\mu} \approx 0$ LogL= -15.925	$\hat{\lambda}$=0.1622 LogL= -15.925	$\hat{\lambda}$=0.6007 $\hat{\alpha}$=0.2168 LogL= -12.775	$\hat{\lambda}$=0.7403 $\hat{\alpha}$=0.2181 $\hat{\mu}$=0.0645 $\hat{\beta}$=0.5114 LogL=-12.543	$\hat{\lambda}$=0.1622 $\hat{\mu} \approx 0$ $\hat{\beta}$=0.8368 LogL= -15.925	$\hat{\lambda}$=0.7659 $\hat{\alpha}$=0.2235 $\hat{\mu}$=0.0626 LogL= -12.775
15. Tangara N=49 n=42	$\hat{\lambda}$=0.2117 $\hat{\mu} \approx 0$ LogL= 5.789	$\hat{\lambda}$=0.2117 LogL= 5.789	$\hat{\lambda}$=1.164 $\hat{\alpha}$=0.2816 LogL= 16.342	$\hat{\lambda}$=1.164 $\hat{\alpha}$=0.2816 $\hat{\mu} \approx 0$ $\hat{\beta}$=0.5724 LogL=16.342	$\hat{\lambda}$=0.2117 $\hat{\mu} \approx 0$ $\hat{\beta}$=2.148 LogL= 5.789	$\hat{\lambda}$=1.164 $\hat{\alpha}$=0.2816 $\hat{\mu} \approx 0$ LogL= 16.342
16. Trogons N=39 n=29	$\hat{\lambda}$=0.1009 $\hat{\mu} \approx 0$ LogL= -29.777	$\hat{\lambda}$=0.1009 LogL= -29.777	$\hat{\lambda}$=0.2662 $\hat{\alpha}$=0.0631 LogL= -27.174	$\hat{\lambda}$=0.2662 $\hat{\alpha}$=0.0631 $\hat{\mu} \approx 0$ $\hat{\beta}$=0.0210 LogL=-27.175	$\hat{\lambda}$=0.1009 $\hat{\mu} \approx 0$ $\hat{\beta}$=1.308 LogL= -29.777	$\hat{\lambda}$=0.2662 $\hat{\alpha}$=0.0631 $\hat{\mu} \approx 0$ LogL= -27.175
17. *Turdus* and allies N=70 n=60	$\hat{\lambda}$=0.2159 $\hat{\mu} \approx 0$ LogL= 29.861	$\hat{\lambda}$=0.2159 LogL= 29.861	$\hat{\lambda}$=0.9237 $\hat{\alpha}$=0.1418 LogL= 36.358	$\hat{\lambda}$=0.9237 $\hat{\alpha}$=0.1418 $\hat{\mu} \approx 0$ $\hat{\beta}$=0.0996 LogL=36.358	$\hat{\lambda}$=0.2159 $\hat{\mu} \approx 0$ $\hat{\beta}$=2.804 LogL= 29.861	$\hat{\lambda}$=0.9238 $\hat{\alpha}$=0.1418 $\hat{\mu} \approx 0$ LogL= 36.358
18. Wrens N=74 n=50	$\hat{\lambda}$=0.1891 $\hat{\mu} \approx 0$ LogL= 1.404	$\hat{\lambda}$=0.1891 LogL= 1.404	$\hat{\lambda}$=1.152 $\hat{\alpha}$=0.2664 LogL= 18.858	$\hat{\lambda}$=1.152 $\hat{\alpha}$=0.2664 $\hat{\mu}$=0.1427 $\hat{\beta} \approx 0$ LogL=18.858	$\hat{\lambda}$=0.1891 $\hat{\mu} \approx 0$ $\hat{\beta}$=2.771 LogL= 1.404	$\hat{\lambda}$=1.152 $\hat{\alpha}$=0.2664 $\hat{\mu} \approx 0$ LogL= 18.858
19. Tringa N=13 n=12	$\hat{\lambda}$=0.1037 $\hat{\mu} \approx 0$ LogL= -18.932	$\hat{\lambda}$=0.1037 LogL= -18.932	$\hat{\lambda}$=0.6012 $\hat{\alpha}$=0.2155 LogL= -14.727	$\hat{\lambda}$=0.6012 $\hat{\alpha}$=0.2155 $\hat{\mu} \approx 0$ $\hat{\beta}$=0.8426 LogL= -14.727	$\hat{\lambda}$=0.1037 $\hat{\mu} \approx 0$ $\hat{\beta}$=1.725 LogL= -18.932	$\hat{\lambda}$=0.6012 $\hat{\alpha}$=0.2155 $\hat{\mu} \approx 0$ LogL= -14.727
20. Meliphaga N=13 n=12	$\hat{\lambda}$=0.1181 $\hat{\mu} \approx 0$ LogL= -17.513	$\hat{\lambda}$=0.1181 LogL= -17.513	$\hat{\lambda}$=0.4465 $\hat{\alpha}$=0.1404 LogL= -15.443	$\hat{\lambda}$=0.4465 $\hat{\alpha}$=0.1404 $\hat{\mu} \approx 0$ $\hat{\beta}$=0.3098 LogL= -15.443	$\hat{\lambda}$=0.1181 $\hat{\mu} \approx 0$ $\hat{\beta}$=2.732 LogL= -17.513	$\hat{\lambda}$=0.4465 $\hat{\alpha}$=0.1404 $\hat{\mu} \approx 0$ LogL= -15.443

C future work

We know how many species live at the present time T and therefore we should condition on that knowledge. We seek

$$f(t_2,t_3,...,t_N \mid T,N) = \frac{f(t_2,t_3,...,t_N \mid T)}{P^4\left(n(t=T)=N\right)}.$$

The denominator is the probability that we have N species at time T in the reconstructed process

$$P^4\left(n(t=T)=N\right) = \left(1-u_T\frac{P(0,T)}{P(0,T)}\right)\left(u_T\frac{P(0,T)}{P(0,T)}\right)^{N-1} = \left(1-u_T\right)u_T^{\ N-1}.$$

And so the density of the speciation times given the number of species today and the age of the tree is

$$f(t_2,t_3,...,t_N \mid T,N) = \frac{f(t_2,t_3,...,t_N \mid T)}{\left(1-u_T\right)u_T^{\ N-1}}.$$

If, we let the process start at the origin with one single lineage, we have an additional parameter which is the time of the origin (or the time of the whole process). Trying to estimate this parameter along with the speciation and extinction rates did not give good results but it can be integrated out as nuisance instead. Since commonly we do not know anything about the time of the origin and are not particular interested in it, we advocate in favour of integrating out the time of the origin. This will yield $f(t_2,t_3,...,t_N \mid N)$ instead. To achieve this we use Gernhard's equation 4 (Gernhard, 2008) which states that $f(s_i \mid t_2 = t) = f(s_i \mid t_{or} = t)$ where s_i is the speciation times unordered, the speciation times are i.i.d, and t_{or} is the time of origin. Gernhard (2008) defines the present time as zero and let time go backwards and therefore the time of the origin t_{or} is the same as the total age of the tree, our T, and her times of speciation is the same as our x_i. In our case Gernhard's equation 4 in is the same as $f(s_i \mid x_2 = t) = f(s_i \mid T = t)$. It is then with some calculations easily confirmed that the density of the speciation times conditioned on T and N for constant rates is the same as

$$f(t_2,t_3,...,t_N \mid T=t,N) = f(t_2,t_3,...,t_N \mid x_2=t,N)$$

$$= (N-1)!\left(\prod_{i=2}^{N}\frac{(\lambda\text{-}\mu)^2\exp\left(-(\lambda\text{-}\mu)x_i\right)}{\left(\lambda-\mu\exp\left(-(\lambda\text{-}\mu)x_i\right)\right)^2}\frac{\lambda-\mu\exp\left(-(\lambda\text{-}\mu)t\right)}{1-\exp\left(-(\lambda\text{-}\mu)t\right)}\right).$$

This is the same density function as on top of page 20 in Gernhard's paper (2008). Theorem 3.2 in (Gernhard, 2008) gives the density function of the age of the tree, $T(t_{or})$, conditioned on that the tree has N species,

$$P(T \mid N) = n\lambda^N (\lambda - \mu)^2 \frac{\left(1 - \exp\left(-(\lambda - \mu)T\right)\right)^{N-1} \exp\left(-(\lambda - \mu)T\right)}{\left(\lambda - \mu\exp\left(-(\lambda - \mu)T\right)\right)^{N+1}}.$$

Through the law of total probability we then have the integral below which Gernhard solves in the same article (2008)

$$f(t_2, t_3, ..., t_N \mid N) = \int_{x_2}^{\infty} f(t_2, t_3, ..., t_N \mid T, N) P(T \mid N) dT =$$

$$= N! \lambda^{N-1} \frac{(\lambda\text{-}\mu)\exp\left(-(\lambda\text{-}\mu)x_2\right)}{\lambda\text{-}\mu\exp\left(-(\lambda\text{-}\mu)x_2\right)} \prod_{i=2}^{N} \frac{(\lambda\text{-}\mu)^2 \exp\left(-(\lambda\text{-}\mu)x_i\right)}{\left(\lambda\text{-}\mu\exp\left(-(\lambda\text{-}\mu)x_i\right)\right)^2} =$$

$$= N! \lambda^{N-1} P(0, x_2) r(x_2) \prod_{i=2}^{N} P(t_i, x_2)^2 r(x_i).$$

This last equation would be our likelihood function for constant rates. If we want to do this for all our models we would need to do the same but for $\lambda(t)$ and $\mu(t)$. We can not just simply exchange all λ into $\lambda(t)$. We would have to find the density of time T given N species but for varying rates and then integrate out T.