



Stockholms
universitet

Birth and Death Models with Time Dependent Rates Applied to Virus Phylogenies

Disa Hansson

Masteruppsats 2013:2
Matematisk statistik
Juni 2013

www.math.su.se

Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm

Birth and Death Models with Time Dependent Rates Applied to Virus Phylogenies

Disa Hansson*

June 2013

Abstract

In this thesis we use a modified birth and death process to model phylogenetic trees reconstructed from sequence data that has been sampled at different times. This is common in organisms where major evolutionary changes can occur in a short period of time, such as viruses, and therefore the dating of the sampling is of great importance. Our main focus is on birth and death processes with time dependent rates, instead of previously used constant rates. Here we apply our model on eight phylogenetic trees reconstructed from different genome segments of influenza A virus. The evolution of influenza virus is affected by the interplay with our immune system, which makes it possible to infer epidemiological parameters from its sequence data. For example the effective reproductive number over time, which is the average number that one infected individual will infect in a population consisting of not only susceptible individuals. It is known that influenza outbreaks have a seasonal behaviour. Our aim with this thesis is to investigate if this seasonal behaviour can be detected using birth and death processes with periodic rate functions on the phylogenetic trees of the influenza A virus. Furthermore, we also want to infer the effective reproductive number. As a main result, models with periodic rate functions have a better fit to the data than constant rate models.

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

Acknowledgements

This paper constitutes a master thesis of 30 ECTS in Biostatistics at the Department of Mathematics at Stockholm University.

I want to thank my supervisor Sebastian Höhna for the help he has given me to write this thesis. I also want to thank my examiner Pieter Trapman for comments and input. Last but not least I want to thank Kaj Börjesson for helpful discussions.

Contents

1 Introduction	2
1.1 Outline	3
2 Background	3
2.1 The reconstructed evolutionary process	4
2.1.1 Homochronous data	4
2.1.2 Heterochronous data	11
3 Method	18
3.1 Periodic Rate functions	18
3.2 Simulations and validation	20
3.2.1 Likelihood surface	20
3.2.2 Simulated data inferred with constant rates	21
3.2.3 Simulated data inferred with periodic speciation rate	24
3.2.4 Differential equations	28
3.3 Inference	32
4 Analysis	33
4.1 Data	33
4.2 Results	34
5 Discussion	37
5.1 Reproductive number	37
5.2 Yearly Period	38
5.3 Studied models and possible drawbacks	38
6 Summary and future work	40
6.1 Summary and conclusions	40
6.2 Future work	40
A Equality between Equation 2 and Thompson (1975)	44
B Complete Results	46
C Simulations	61
D R code	67

1 Introduction

Humans worldwide get infected by influenza A virus every year. The interplay between our immune system and the virus's ability to adapt affect how the virus evolves. Here we will use sequence data and molecular phylogenetics to address epidemiological questions.

The process of evolution is often represented by a phylogenetic tree. A phylogenetic tree is a binary tree where the ancestral node represents the most recent common ancestor (MRCA) of all other lineages. A split in the tree represents a speciation event, and a termination an extinction event. Molecular sequences are used to build the evolutionary history. These are usually obtained at different time points, called heterochronous sequencing. In larger organisms, such as mammals, this is of minor relevance since little evolution occurs between the years and we can assume the sequencing was done at the same time point, called homochronous sequencing. However, in viruses many mutations can occur between samplings, and we should take into account that the data was gathered at different time points. When performing inference on virus evolution we should not use models that assume that sequencing has been done at the same time. When inferring the substitution rate, the rate of genetic diversification, Twiddy et al. (2003) showed that models taking into account the different datings, when applied to Dengue virus sequences fitted much better than models not doing so. This is not only needed in models for fast evolving organisms but also for the incorporation of fossils, since methods of sequencing fossil DNA and especially mtDNA has become more reliable (Orlando et al. 2002). Therefore we need models that incorporate fossil data, as has recently been done for morphological data (Ronquist et al. 2012).

One can use phylogenetics to answer questions usually asked by epidemiologists (Pybus et al. 2001; Stadler et al. 2013; Leventhal et al. 2012). For example to estimate the basic reproductive number R_0 , which is equal to the expected number of infections that one infected individual will cause when all other individuals in the population are susceptible. This is mainly done by using a coalescent model (Volz et al. 2009; Pybus et al. 2001). However, the coalescent approach is not suitable for this kind of processes, but instead for population dynamics and it has some major drawbacks (Stadler et al. 2012). It has been widely used because of its simplicity and mathematical tractability. We will instead try to incorporate heterochronous sampling using the birth and death processes. Under the birth and death process speciation, extinction and sampling rates needs to be specified. Heterochronous sampling using birth and death processes has been done by Stadler et al. (2012) but only for constant rates and in Stadler et al. (2013) with piecewise constant rates. Only using constant rates for organism evolution is not realistic and as stated several times, we should use rates as functions of time instead. If time dependent rates are

used we can for example allow periods where the extinction rate is higher than the speciation rate. This is important when trying to model the spread of an infection; we can capture more than the exponential increase of infected individuals in the beginning of an epidemic. We can now also capture the saturation phase when the number of susceptibles decreases.

In this thesis we will derive the likelihood for heterochronous data and time dependent rates. We will use this likelihood to perform inference on eight phylogenetic trees reconstructed from different genome segments of Influenza A virus. We especially want to infer the reproductive number over time. As we know, the influenza infections come in clearly periodic intervals and we are interested if we can detect this behaviour from the phylogenetic trees.

1.1 Outline

The structure of the thesis is as follows: In Section 2 the background of the thesis is presented. First we will define the birth and death process used to model phylogenetic trees reconstructed from homochronous data. The focus will be put on the derivations for time dependent rates. Next, the probability density of a phylogenetic tree reconstructed from heterochronous data will be derived. This expression will be used to make inference of the evolutionary history of influenza A virus.

In Section 3 we will go through the different rate functions used to answer our main question; 'Can we detect the seasonal behaviour of Influenza?'. In this section we will also try to validate if our likelihood gives reliable results by simulations.

In Section 4 we will go through the data used in this study and the results obtained from our analysis.

2 Background

To model the splitting of lineages we will use birth and death processes. The lineages can represent different things, such as individuals, populations or species. We assume that we start with one lineage at $t = 0$. At any time $t > 0$, a lineage can split into two lineages with rate $\lambda(t)$ or go extinct with rate $\mu(t)$. Thus we have the same speciation rate for every lineage at a given time and the same applies to the extinction rate. In a short time period h three outcomes are possible: a birth event, a death event or no event. If $n(t)$ is the number of lineages at time t this process is described by

$$n(t+h) = \begin{cases} n(t) + 1 & \text{w.p. } \lambda(t)n(t)h \\ n(t) - 1 & \text{w.p. } \mu(t)n(t)h \\ n(t) & \text{w.p. } 1 - (\lambda(t) + \mu(t))n(t)h \end{cases}$$

Most analysis on phylogenies have been on trees based solely on extant species/genes. In case of virus evolution we can observe the extinction of branches. The derivation of the likelihood for heterochronous data will be done as follows: we will describe the underlying process which results in the species diversity with homochronous data in 2.1.1 and then how to expand this to heterochronous data in section 2.1.2.

2.1 The reconstructed evolutionary process

The true tree of species evolutionary relationship is a tree where all species is represented, both extinct and extant species. This can be represented by an ordinary birth and death process where we can observe births of new lineages as well as deaths, terminations of lineages. Figure 1 a) is an illustration of such a process. To build a phylogeny we reconstruct it from molecular or morphological data. In our framework a reconstructed process is a birth and death process where we have pruned off all lineages that we do not have any data from as seen in Figure 1 b). If we allow for each lineage to have a probability of being sampled we can take into account incomplete taxon sampling, Figure 1 c).

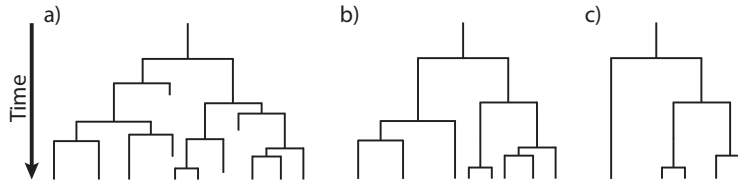


Figure 1: a) is an example of a complete phylogeny where extinct unsampled lineages are kept. b) A reconstructed phylogeny where only sampled lineages are kept. c) A reconstructed phylogeny with incomplete taxon sampling, each lineage has the same probability of being sampled

In Hansson (2011) we derived the likelihood for a reconstructed phylogeny from homochronous data with time dependent rates and incomplete taxon sampling. This was also done in Morlon et al. (2011) but using a different approach.

2.1.1 Homochronous data

Let us now consider in more detail how the phylogenetic tree is usually modelled. At a given time point a split occurs with the same probability and independently on every existing lineage. We only consider splits that results in both the lineages surviving until present time T . If we have a phylogeny with n surviving lineages we will have $n - 1$ splitting events where both lineages survived until present time.

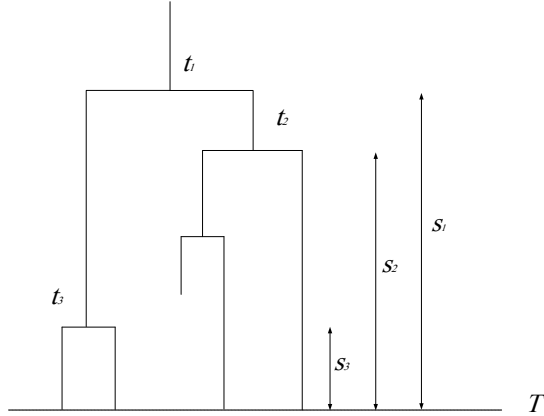


Figure 2: Tree with 4 surviving lineages. t_i denotes the splitting times and s_i denotes the time from each splitting event until present time T

We want to derive the probability density for the process generating reconstructed phylogenetic trees. To achieve this we will first give the number of different trees that can be created from a tree with n labelled tips, i.e. fixed species. Secondly, some necessary probabilities and functions will be presented. Thirdly, a lemma needed to simplify one part in the probability density is presented and proven for time dependent rates. Before, this lemma has only been proven for constant rates.

The number of phylogenetic trees with n species can be found in the following way. First consider a tree with labelled nodes and unlabelled tips, called an "unlabelled tree". The first splitting event can only happen at the root, the second splitting event can happen on two different branches, the third splitting event can happen on three different branches and the final splitting event on any of the $n - 1$ branches. Thus, there is $(n - 1)!$ different "unlabelled trees". If now the tips are labelled as well they can be placed on n different branches generating $n!$ different ways to place the labels. However, each of these ways will for the $n - 1$ splitting events count one tree twice. Let us study Figure 2 and the split at time t_3 . If we are to place two species labels on these two final lineages we can place them in two ways, but these two ways generate the same phylogenetic tree. This gives us the total number of trees we can create from n species and a labelled history to be

$$\frac{n!(n - 1)!}{2^{n-1}}$$

where

$$n!/2^{n-1} \tag{1}$$

is the number of different distinct labellings of the tips of an unlabelled history (Edwards 1970). The phylogenies does only have labelled tips and the orientation within is not fixed, in other words, the phylogenies we want to infer have tips with specified labels but do not distinguish between left and right. Thus, when deriving the density for a phylogenetic tree it is needed to divide the final expression with Equation (1) (Thompson 1975; Stadler et al. 2013).

Let us now go through notations, probabilities and some equations needed. Let N be the number of lineages at present, let \vec{t} denote the set of splitting events $\{t_1, \dots, t_{N-1}\}$ and let τ denote the relation between the lineages, a topology describing the phylogenetic tree. We seek the density $f(\vec{t}, N, \tau; \lambda(t), \mu(t))$. At present time T each lineage has a probability ρ of being sampled. To model this a 'mass extinction' at the present time will be imposed where each lineage survives with a probability ρ . This will change the extinction rate to be

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

where $\delta(t)$ is the dirac delta function. Let $P_N(0, t)$ denote the probability of having N species at time t if the process started at time 0 with one species. Also let $P_S(t, T)$ denote the probability that the process has not gone extinct from time t to time T , i.e. at least one lineage survived to time T . We then have from Kendall (1949)

$$\begin{aligned} P_0(0, t) &= 1 - P_S(0, t) \\ P_1(0, t) &= P_S(0, t)^2 e^{r(0, t)} \\ P_N(0, t) &= P_S(0, t)^2 e^{r(0, t)} (1 - P_S(0, t) e^{r(0, t)})^{N-1} \end{aligned} \quad (2)$$

where

$$r(t, \tau) = \begin{cases} \int_t^\tau \mu(s) - \lambda(s) ds & \text{if } \tau < T \\ \int_t^\tau \mu(s) - \lambda(s) ds - \ln \rho & \text{if } \tau \geq T \end{cases} \quad (3)$$

and

$$P_S(t, T) = \frac{1}{1 + \int_t^T \mu(s) \exp(r(t, s)) ds}. \quad (4)$$

If T is the present time

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

With constant rates $\lambda(t) = \lambda$ and $\mu(t) = \mu$ and complete taxon sampling, Equation 2 is the same as in Thompson (1975). This is shown in Appendix A.

Let us show this next lemma to be able to simplify one part in the derivation of the density $f(\vec{t}, N, \tau; \lambda(t), \mu(t))$. Thompson (1975) states why it must hold.

Lemma 1. *If we start with one species at time 0 and want to end up with one species at time $t > 0$ we can have k species at an earlier time point $t - s$, $0 < s < t$, one of these k species must have one descendant at time t and the remaining $k - 1$ must have gone extinct. Summing over all possible values of k gives*

$$\sum_{k=1}^{\infty} k P_0(t-s, t)^{k-1} P_1(t-s, t) P_k(0, t-s) = P_1(0, t)$$

Proof. Let us begin by using Equation (2) on $P_k(0, t-s)$

$$\begin{aligned} & \sum_{k=1}^{\infty} k P_0(t-s, t)^{k-1} P_1(t-s, t) P_k(0, t-s) \\ &= \sum_{k=1}^{\infty} k P_0(t-s, t)^{k-1} P_1(t-s, t) P_S(0, t-s)^2 e^{r(0, t-s)} (1 - P_S(0, t-s) e^{r(0, t-s)})^{k-1} \\ &= P_1(t-s, t) P_S(0, t-s)^2 e^{r(0, t-s)} \\ & \quad \times \sum_{k=1}^{\infty} k (P_0(t-s, t) (1 - P_S(0, t-s) e^{r(0, t-s)}))^{k-1} \end{aligned} \tag{5}$$

Now consider first the expression outside the summation in Equation (5), which is equal to

$$\begin{aligned} & P_S(t-s, t)^2 e^{r(t-s, t)} P_S(0, t-s)^2 e^{r(0, t-s)} \\ &= (P_S(t-s, t) P_S(0, t-s))^2 e^{r(0, t)} \end{aligned} \tag{6}$$

The summation in Equation (5) can be simplified by noting that if we assume that the birth rate is overall larger than the death rate:

$\int_0^{t-s} \lambda(x) dx > \int_0^{t-s} \mu(x) dx$ then $P_0(t-s, t) (1 - P_S(0, t-s) e^{r(0, t-s)}) < 1$. We can then use that if $|x| < 1$ then $\sum_{k=1}^{\infty} k x^{k-1} = \frac{1}{(1-x)^2}$.

$$\begin{aligned}
& \sum_{k=1}^{\infty} k(P_0(t-s, t)(1 - P_S(0, t-s)e^{r(0, t-s)}))^{k-1} \\
&= (1 - P_0(t-s, t)(1 - P_S(0, t-s)e^{r(0, t-s)}))^{-2} \\
&= (1 - (1 - P_S(t-s, t))(1 - P_S(0, t-s)e^{r(0, t-s)}))^{-2} \tag{7} \\
&= (P_S(0, t-s)e^{r(0, t-s)} + P_S(t-s, t) - P_S(t-s, t)P_S(0, t-s)e^{r(0, t-s)})^{-2} \\
&= (P_S(0, t-s)e^{r(0, t-s)}(1 - P_S(t-s, t)) + P_S(t-s, t))^{-2}
\end{aligned}$$

If we now combine (6) and (7) we get

$$\begin{aligned}
& \left(\frac{P_S(t-s, t)P_S(0, t-s)}{P_S(0, t-s)e^{r(0, t-s)}(1 - P_S(t-s, t)) + P_S(t-s, t)} \right)^2 e^{r(0, t)} \\
&= \left(\frac{P_S(0, t-s)}{P_S(0, t-s)e^{r(0, t-s)} \frac{(1 - P_S(t-s, t))}{P_S(t-s, t)} + 1} \right)^2 e^{r(0, t)} \\
&= \left(\frac{P_S(0, t-s)}{P_S(0, t-s)e^{r(0, t-s)} \left(\frac{1}{P_S(t-s, t)} - 1 \right) + 1} \right)^2 e^{r(0, t)} \\
&= \left(\frac{1}{e^{r(0, t-s)} \left(\frac{1}{P_S(t-s, t)} - 1 \right) + \frac{1}{P_S(0, t-s)}} \right)^2 e^{r(0, t)}
\end{aligned}$$

Now it is only a matter of using Equation (4) to obtain the wanted result

$$\begin{aligned}
&= \frac{1}{\left(e^{r(0,t-s)} \int_{t-s}^t \mu(x) e^{r(t-s,x)} dx + 1 + \int_0^{t-s} \mu(x) e^{r(0,x)} dx \right)} 2 e^{r(0,t)} \\
&= \frac{1}{\left(\int_{t-s}^t \mu(x) e^{r(0,x)} dx + 1 + \int_0^{t-s} \mu(x) e^{r(0,x)} dx \right)} 2 e^{r(0,t)} \\
&= \frac{1}{\left(1 + \int_0^t \mu(x) e^{r(0,x)} dx \right)} 2 e^{r(0,t)} = P_S(0, t)^2 e^{r(0,t)} = P_1(0, t).
\end{aligned}$$

□

In the following theorem the probability density of a reconstructed tree will be given. Let $\mathbb{1}$ be the indicator function and let

$$\mathbb{1}(t_i, t_j) := \mathbb{1}(\text{branch joining the split at } t_i \text{ and split at } t_j)$$

where $t_i < t_j$. Also, let k_{ij} be the number of offsprings from the lineage at t_i until time t_j .

Theorem 2. *The density of a reconstructed phylogenetic tree is given by*

$$f(\vec{t}, N, \tau; \lambda(t), \mu(t)) = \frac{2^{N-1}}{N!} P_1(t_1, T)^2 \prod_{j=2}^{N-1} \lambda(t_j) P_1(t_j, T) \quad (8)$$

Proof. Splitting of lineages are independent events and thus the sought probability is just the product of the probability of the branches where $\mathbb{1}(t_i, t_j) = 1$ for $i = 1, \dots, N-1$ and $j = 2, \dots, N$. $k_{ij} > 0$ or $k_{iN} = 1$. If we have two splitting times t_i and t_j joined by a branch, $\mathbb{1}(t_i, t_j) = 1$, we will with probability

$$P_{k_{ij}}(t_i, t_j)$$

have k_{ij} offsprings from time t_i until t_j . At time t_j one of these lineages will split with probability

$$\lambda(t_j) k_{ij}.$$

The rest $k_{ij} - 1$ lineages must go extinct before time T , they do so with probability

$$P_0(t_j, T)^{k_{ij}-1}.$$

Even if one of the $k_{ij} - 1$ lineages split in a later time than t_j and then go extinct this new lineage will not be considered since it has no ancestor in our tree. We also know that all branches between a splitting time t_s and $t_N = T$ must have $k_{sN} = 1$, since these are the final branches before time T . We also need to divide by the number of distinct labellings of the unlabelled history. Note that t_N is not a splitting event but the present time T .

$$f(\vec{t}, N, \tau; \lambda(t), \mu(t)) = \frac{2^{N-1}}{N!} \prod_{\substack{\mathbb{1}(t_i, t_j)=1 \\ 1 \leq i \leq N-1 \\ 2 \leq j \leq N}} \lambda(t_j) \left(\sum_{k_{ij} > 0, k_{iN}=1} k_{ij} P_0(t_j, T)^{k_{ij}-1} P_{k_{ij}}(t_i, t_j) \right)$$

We note that the summation is almost the same as in Lemma 1 apart from the term $P_1(t_j, T)$. The second summation index, $k_{iN}=1$, contributes in fact with only one term, namely $P_1(t_i, T)$. Also, the last branches before present time will not contribute with any $\lambda(t_N)$ since there is no splitting event at time T . If we multiply the nominator and denominator with $P_1(t_j, T)$ we get

$$\begin{aligned} f(\vec{t}, N, \tau; \lambda(t), \mu(t)) &= \frac{2^{N-1}}{N!} \prod_{j=2}^{N-1} \lambda(t_j) \prod_{\substack{\mathbb{1}(t_i, t_j)=1 \\ \text{inner branch}}} \frac{P_1(t_i, T)}{P_1(t_j, T)} \prod_{\substack{\mathbb{1}(t_i, t_N)=1 \\ \text{outer branch}}} P_1(t_i, T) \\ &= \frac{2^{N-1}}{N!} \prod_{j=2}^{N-1} \lambda(t_j) \prod_{\substack{\mathbb{1}(t_i, t_j)=1 \\ 1 \leq i \leq N-1 \\ 2 \leq j \leq N}} \frac{P_1(t_i, T)}{P_1(t_j, T)} \end{aligned}$$

This product can be simplified by studying a specific tree (Figure 2). From one splitting time t_i there are two branches leading to, say, t_{i1} and t_{i2} , the contribution from t_i to our probability is

$$\frac{P_1(t_i, T)}{P_1(t_{i1}, T)} \frac{P_1(t_i, T)}{P_1(t_{i2}, T)}.$$

From this we can conclude that $P_1(t_1, T)$ will be multiplied twice since t_1 has two branches and $P_1(t_{i1}, T)$ or $P_1(t_{i2}, T)$ will never become $P_1(t_1, T)$, therefore this term is left alone. However, $P_1(t_2, T)$ will also be multiplied twice but once divided with. That is, each splitting event is once an ending point and twice a starting point, except the very first one which is only a starting point. $P_1(t_3, T)$ will also be multiplied twice and then divided once with, and so on until $P(t_N, T)$ which is simply equal to 1. Thus we have our result

$$f(\vec{t}, N, \tau; \lambda(t), \mu(t)) = \frac{2^{N-1}}{N!} P_1(t_1, T)^2 \prod_{j=2}^{N-1} \lambda(t_j) P_1(t_j, T)$$

□

This likelihood function is the same as in Nee et al. (1994) except that they condition on the survival of the two lineages from the first splitting event. Their likelihood is this one divided with $P_S(t_1, T)^2$. If we have no information of the length of the lineage until the first splitting event we should begin the process at the first splitting event t_1 and thus condition on the survival of the first two branches.

2.1.2 Heterochronous data

Now we will change the underlying model to fit the data that has been sampled throughout time. We will assume that a sampled lineage cannot be sampled again. If we want to model disease transmission it is more realistic to assume that a sampled lineage is terminated, i.e. becoming non-infectious. A sampling event in disease modelling is equivalent to that the infected human is becoming aware of the disease and either gets treatment or changes his/her behaviour. Especially since a sampling from our virus data is a confirmed case of influenza this should result in the infected individual becoming more careful. If we instead only want to model virus evolution and for example studied the HIV virus it would make sense to allow for more than one sample per lineage. However, if we are interested in the spreading of the disease we would argue that a patient becoming aware of having HIV would become more careful and then a sampling is equivalent to becoming non-infectious, even if the virus can evolve inside this patient.

The process proceeds as follows, we start at time $t = 0$ with one lineage, at a given time before time T a lineage can give birth (transmission event), die or become sampled. A sampling before present time is in this model equivalent to extinction since a sampled lineage stops. Lineages surviving to the present will be sampled with a probability ρ . Additional to this process compared to the one in the previous section is a sampling rate which we will denote $\psi(t)$. If $\psi(t) = 0$ the process is reduced to the process explained in the previous section. All lineages not being sampled will be deleted since we want to model a reconstructed phylogeny. Figure 3 shows an example of a tree from a birth and death process where we can see sampled and non-sampled lineages. In this model we do not only want to consider splitting events where both the new lineages survived to present time, but also the incorporation of sampled lineages in the past. In this process three things can happen in a short time period h , a new lineage is born (a birth event), a lineage is removed (a death or sampling event occur) or no event

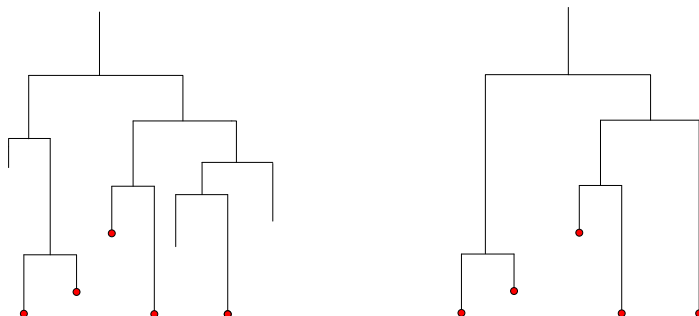


Figure 3: The left tree illustrates a complete tree with 3 surviving lineages and 5 extinct lineages. The right tree illustrates a tree under heterochronous sampling with 3 surviving lineages and 2 extinct sampled lineages

$$n(t+h) = \begin{cases} n(t) + 1 & \text{w.p. } \lambda(t)n(t)h \\ n(t) - 1 & \text{w.p. } (\mu(t) + \psi(t))n(t)h \\ n(t) & \text{w.p. } 1 - (\lambda(t) + \mu(t) + \psi(t))n(t)h \end{cases}$$

The derivation of the likelihood will be done in the same procedure as Stadler (2010). First we will define the probabilities needed to derive the likelihood. These probabilities will be obtained by differential equations. Then, the likelihood will be derived.

The two probabilities needed to derive the likelihood are the following:

$P_0^S(t, T)$, the probability that a lineage alive at time t before present time has no sampled extinct or extant descendants.

$P_1^S(t, T)$, the probability that a lineage alive at time t before present time has one sampled extant and no sampled extinct descendant.

We have used the superscript S for 'sampling' to distinguish the probabilities in the previous section. We know that $P_0^S(T, T) = 1 - \rho$ and that $P_1^S(T, T) = \rho$. Let us begin with how to get $P_0^S(t, T)$. For a lineage alive at time t before present time not to be sampled until time T it needs either to die, remain not being sampled or it can give birth to another lineage and both the lineages do not get sampled. If we move a short time step h from t this would yield

$$P_0^S(t+h, T) = \mu(t)h + (1 - (\lambda(t) + \mu(t) + \psi(t))h)P_0^S(t, T) + \lambda(t)P_0^S(t, T)^2h$$

\Leftrightarrow

$$\frac{P_0^S(t+h, T) - P_0^S(t, T)}{h} = \mu(t) - (\lambda(t) + \mu(t) + \psi(t))P_0^S(t, T) + \lambda(t)P_0^S(t, T)^2$$

When $h \rightarrow 0$ we have the following differential equation

$$\frac{d}{dt}P_0^S(t, T) = \mu(t) - (\lambda(t) + \mu(t) + \psi(t))P_0^S(t, T) + \lambda(t)P_0^S(t, T)^2. \quad (9)$$

In a similar fashion we obtain a differential equation for $P_1^S(t, T)$. For a lineage alive at time t to have exactly one sampled extant lineage at present time either nothing happens and the lineage survives or a birth happens and either of the two lineages does not get sampled ($2P_0^S(t, T)$) and the other one is sampled at present time.

$$\frac{d}{dt}P_1^S(t, T) = -(\lambda(t) + \mu(t) + \psi(t))P_1^S(t, T) + 2\lambda(t)P_0^S(t, T)P_1^S(t, T) \quad (10)$$

Note, that if we instead wanted to calculate the probability for a lineage alive at time t of having exactly one sampled lineage at an earlier time than present time $s < T$ the differential equation would look like

$$\frac{d}{dt}P_1^S(t, s) = -(\lambda(t) + \mu(t) + \psi(t))P_1^S(t, s) + 2\lambda(t)P_0^S(t, T)P_1^S(t, s)$$

Since, to have exactly one sampled lineage at time s the lineage at time t undergo no change (no splitting or no termination) and then get sampled at time s or it is splitted into two lineages and one of them get sampled at time s and the other one doesn't get sampled during the whole time $P_0^S(t, T)$. But then we do not have an initial condition. The solution to these differential equations, Equation 9 and 10, for constant rates is given in Stadler (2010). For time dependent rates we will need to solve the equations for the different rates. To solve the two cases below we used Mathematica.

- if $\mu(t) = \psi(t) = 0$

The solution to Equation (9) if we do not have sampling in the past or extinction is

$$P_0^S(t, T) = \frac{(1 - \rho)e^{-\int_t^T \lambda(x)dx}}{(1 - \rho)e^{-\int_t^T \lambda(x)dx} + \rho} = 1 - \frac{1}{1 - (\rho - 1)e^{-\int_t^T \lambda(x)dx - \ln(\rho)}}$$

This is easily verified to be exactly the same as $P_0(t, T)$ in Equation (2) (when $\mu(t) = 0$) which is not surprising since we do not have any sampling through time. Also, if $\rho = 0$ then $P_0^S(t, T) = 1$ as expected. If we have no sampling in the past nor today the probability of having no samples should be equal to 1.

- if $\mu(t) = 0$

The solution to Equation (9) if we have sampling through time but no extinction is

$$P_0^S(t, T) = \frac{(1 - \rho)e^{-\int_t^T \lambda(x) + \psi(x) dx}}{1 - (1 - \rho) \int_t^T e^{-\int_t^x \lambda(s) + \psi(s) ds} \lambda(x) dx}$$

If we plug in constant rates λ and ψ this will become exactly the same solution as in Stadler (2010) Equation (1) when $\mu = 0$. If we want to solve the differential equation for general $\lambda(t)$, $\mu(t)$, and $\psi(t)$ we cannot find a closed expression and have to use numerically methods.

The solution to Equation 10 is

$$P_1^S(t, T) = \rho e^{-\int_t^T \lambda(s) + \mu(s) + \psi(s) - 2\lambda(s) P_0^S(s, T) ds} =: \rho P_1^I(t, T) \quad (11)$$

Where we denote the integral part by $P_1^I(t, T)$

Now we want to derive the density of a phylogenetic tree τ with heterogeneous sampling. We will approach this in the same way as in Stadler (2010). Let $f_l(t)$ be the probability density that a lineage l alive at time t evolved between t and present time T as seen in τ , see Figure 4, then

$$\frac{d}{dt} f_l(t) = -(\lambda(t) + \mu(t) + \psi(t)) f_l(t) + 2\lambda(t) P_0^S(t, T) f_l(t). \quad (12)$$

This is the same kind of differential equation as Equation (10) but with a different initial condition. Let s be the ending time of the lineage l , if sampling corresponds to a termination of a branch we will have the following

$$f_l(s) = \begin{cases} \rho & \text{if } l \text{ has no descendant lineage at } s = T \\ \psi(s) & \text{if } l \text{ has no descendant lineage at } s < T \\ \lambda(s) f_{l_1}(s) f_{l_2}(s) & \text{if } l \text{ has two descendant lineages } l_1 \text{ and } l_2 \end{cases} \quad (13)$$

The solution to 13 will be

$$f_l(t) = \begin{cases} P_1^S(t, s) & \text{if } l \text{ has no descendant lineage at } s = T \\ \frac{\psi(s) P_1^S(t, s)}{\rho} & \text{if } l \text{ has no descendant lineage at } s < T \\ \frac{\lambda(s) f_{l_1}(s) f_{l_2}(s) P_1^S(t, s)}{\rho} & \text{if } l \text{ has two descendant lineages } l_1 \text{ and } l_2 \end{cases}$$

$$= \begin{cases} P_1^S(t, T) & \text{if } l \text{ has no descendant lineage at } s = T \\ \frac{\psi(s)P_1^S(t, T)}{P_1^S(s, T)} & \text{if } l \text{ has no descendant lineage at } s < T \\ \frac{\lambda(s)f_{l_1}(s)f_{l_2}(s)P_1^S(t, T)}{P_1^S(s, T)} & \text{if } l \text{ has two descendant lineages } l_1 \text{ and } l_2 \end{cases} \quad (14)$$

The last step uses that $\frac{P_1^S(t, s)}{\rho} = P_1^I(t, s) = \frac{P_1^I(t, T)}{P_1^I(s, T)} = \frac{P_1^S(t, T)}{P_1^S(s, T)}$.

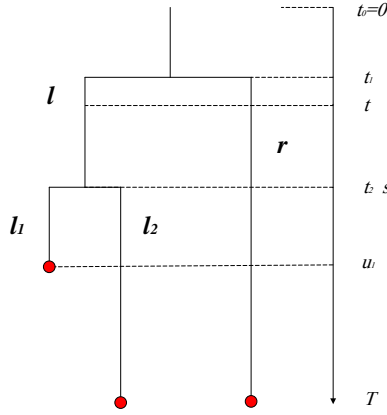


Figure 4: A tree illustrating how to derive the density for a phylogenetic tree with heterochronous sampling. For a lineage l alive at time t the derivation of the density $f_l(t)$ that this lineage will evolve as seen in the tree is done by putting up a differential equation. The initial condition depends on what happens with l . In this figure l split into l_1 and l_2 at time s .

Before, when we had homochronous sampling, the data were the splitting times. Now, when we have heterochronous sampling, additional to the splitting times \vec{t} we will have sampling points. Let m be the number of sampled lineages in the past and let k be the sampled lineages at present time. Let us denote the sampling times with $\vec{u} = \{u_1, \dots, u_{m+k}\}$. We will have $k + m - 1$ splitting events, $\vec{t} = \{t_1, \dots, t_{k+m-1}\}$, let $t_0 = 0$ be the start of the process. Some of these splitting events will be connected to another splitting event, some to sampling events and some to present time. The only difference from Equation (8) is that we also need to consider the splitting events t_i connected to sampling events u_i . To obtain the likelihood for the whole tree we start with the first lineage $f(t_0)$ and recursively work our way through the tree. Let us look at the tree in Figure 4. If we start the process at t_0 then $k = 2$ and $m = 1$. Denote the left branch at t_1 as we have done in the figure with l and the

right branch with r , then

$$\begin{aligned}
f(t_0) &= \frac{\lambda(t_1)f_l(t_1)f_r(t_1)P_1^S(t_0, T)}{P_1^S(t_1, T)} \\
f_r(t_1) &= P_1^S(t_1, T) \\
f_l(t_1) &= \frac{\lambda(t_2)f_{l_1}(t_2)f_{l_2}(t_2)P_1^S(t_1, T)}{P_1^S(t_2, T)} \\
f_{l_1}(t_2) &= \frac{\psi(u_1)P_1^S(t_2, T)}{P_1^S(u_1, T)} \\
f_{l_2}(t_2) &= P_1^S(t_2, T).
\end{aligned}$$

All together we have

$$\begin{aligned}
f(t_0) &= \frac{\lambda(t_1)\lambda(t_2)\psi P_1^S(t_0, T)P_1^S(t_1, T)P_1^S(t_1, T)P_1^S(t_2, T)P_1^S(t_2, T)}{P_1^S(t_1, T)P_1^S(t_2, T)P_1^S(u_1, T)} \\
&= \frac{\lambda(t_1)\lambda(t_2)\psi(u_1)P_1^S(t_0, T)P_1^S(t_1, T)P_1^S(t_2, T)}{P_1^S(u_1, T)}.
\end{aligned}$$

We note as in the proof of Theorem 2 that each splitting event is once an ending point and twice a starting point. All $k + m$ sampling points are ending points. The main difference from homochronous sampling is the m sampling points in the past. So, much like in Theorem 2 but with m additional ending points in the past we get

$$f(\vec{t}, \vec{u}, \tau; \lambda(t), \mu(t), \psi) \propto P_1^S(t_0, T) \prod_{j=1}^{k+m-1} \lambda(t_j)P_1^S(t_j, T) \prod_{j=1}^m \frac{\psi(u_j)}{P_1^S(u_j, T)} \quad (15)$$

Since all trees we do inference on have survived we need to condition on at least one sampled lineage. That is, we need to divide Equation (15) with $1 - P_0^S(0, T)$. The data we are considering here start at the MRCA (most recent common ancestor). We do not know how long the first lineage has lived before the first speciation event occurred, therefore, as in Hansson (2011), we will put the first known speciation event to 0.

In the data we consider here there is only one last lineage in each phylogeny reaching the last sampling time. Therefore it is not as relevant to speak of a probability for that lineage to be sampled at present time and it is better to remove ρ from the model. In this case $\rho = 0$ and the condition to Equation (9) is $P_0^S(T, T) = 1$. This was also done in Stadler et al. (2012). We could argue that when we are considering such fast evolving organisms as viruses, and especially influenza viruses, we cannot have a positive probability of sampling this very moment, but we can instead have a rate of sampling and a probability of being sampled in a time interval. Let m denote the sampling events, we will then have $m - 1$ splitting events. The differential Equation (12) will look the same except for the initial condition.

$$f_l(s) = \begin{cases} \psi(s) & \text{if } l \text{ has no descendant lineage} \\ \lambda(s)f_{l_1}(s)f_{l_2}(s) & \text{if } l \text{ has two descendant lineages } l_1 \text{ and } l_2 \end{cases} \quad (16)$$

The solution to this is

$$f_l(t) = \begin{cases} \frac{\psi(s)P_1^I(t, T)}{P_1^I(s, T)} & \text{if } l \text{ has no descendant lineage} \\ \frac{\lambda(s)f_{l_1}(s)f_{l_2}(s)P_1^I(t, T)}{P_1^I(s, T)} & \text{if } l \text{ has two descendant lineages } l_1 \text{ and } l_2 \end{cases} \quad (17)$$

And our likelihood becomes

$$f(\vec{t}, \vec{u}, \tau; \lambda(t), \mu(t), \psi(t)) \propto P_1^I(t_0, T) \prod_{j=1}^{m-1} \lambda(t_j) P_1^I(t_j, T) \prod_{j=1}^m \frac{\psi(u_j)}{P_1^I(u_j, T)} \quad (18)$$

Note, for Equation (18) to be a density it should be divided by the number of distinct labellings of an unlabelled history, $m!/2^{m-1}$. This is however a constant and is disregarded in the maximization of the likelihood. This will lead to positive log-likelihoods.

3 Method

In this section we will first explain the different rate functions used to do inference on our set of phylogenies. Then we will try to validate how well our models perform with simulations. We will also try to investigate how well the numerical differential equation works and for what parameter values it may not work. Lastly, we will explain how the inference on the empirical data was done.

3.1 Periodic Rate functions

To do inference the rate functions need to be specified. Previously, mainly constant rate models have been used when using birth and death models on heterochronous data (Didier et al. 2012; Stadler et al. 2012). In Section 4.1 we will go through the data used in this study. But first, let us only study the NS1/2 phylogeny in Figure 5. There is clearly a periodic behaviour; once a year there are several new bifurcation events (except in the end of 2000 and the beginning of 2001) and then the lineages dies out (are sampled). This indicates that constant rates for speciation and extinc-

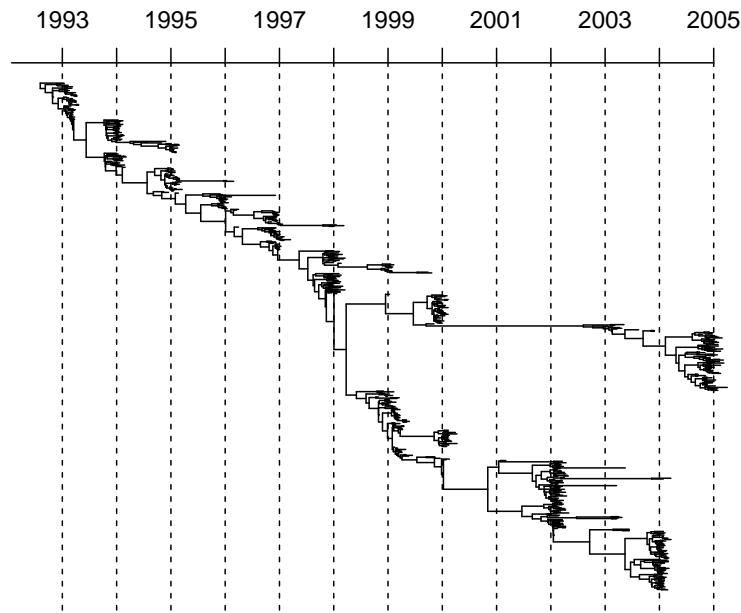


Figure 5: NS1/2 phylogeny. Time is measured in years. There is bursts of speciation events before every new year.

tion rates seem unlikely. Also simple decreasing or increasing functions for the rates does not seem likely. We will instead try to use different periodic functions. We will also assume that the sampling rate is constant even though one could specify a non-constant function as well.

First a periodic speciation rate and a constant extinction and sampling rate will be used. We can then directly use our likelihood function to do inference. We will put $\lambda(t) = \alpha \sin(\beta t) + \lambda_m$. With this function for the speciation rate α is the amplitude, β the period and λ_m is a value larger or equal to α ensuring that the speciation rate is never negativ. λ_m is also the initial speciation rate. If α is 0 we get a constant rate and thus if the underlying data is generated from a constant speciation rate we could expect the estimated speciation rate either to have a very low amplitude (low α) or very long period (low β). By using the sine function we decide that the speciation rate begins at the value λ_m and then increases if the amplitude is positive or decreases if the amplitude is negative. We will also add an extra model with an additional parameter δ to govern the shift of the period, $\lambda(t) = \alpha \sin(\beta t + \delta) + \lambda_m$. Here λ_m is not the initial speciation rate but the value the speciation rate will oscillate around. We will also add models with a periodic extinction rate, $\mu(t) = \alpha \cos(\beta t + \delta) + \mu_m$, again $\alpha \leq \mu_m$ to ensure that the extinction rate is positive. Here μ_m is the value the periodic extinction rate oscillate around. If β , the parameter corresponding to the period, is equal to 1 we would have a period of $2\pi \approx 6.28$ years. Thus, given the influenza phylogenies it is expected to get an estimated β close to 6.28 value to get a yearly period. The models used in this study are listed In Table 1.

Table 1: Different models

Model	$\lambda(t)$	$\mu(t)$	$\psi(t)$
CS	λ	0	ψ
CS & CE	λ	μ	ψ
PS	$\alpha \sin(\beta t) + \lambda_m$	0	ψ
PS & CE	$\alpha \sin(\beta t) + \lambda_m$	μ	ψ
PS shift & CE	$\alpha \sin(\beta t + \delta) + \lambda_m$	μ	ψ
CS & PE	λ	$\alpha \cos(\beta t) + \mu_m$	ψ
CS & PE shift	λ	$\alpha \cos(\beta t + \delta) + \mu_m$	ψ
PS & PE	$\alpha_1 \sin(\beta_1 t) + \lambda_m$	$\alpha_2 \cos(\beta_2 t) + \mu_m$	ψ

CS stands for constant speciation, CE constant etinction, PS periodic speciation, PE periodic extinction

3.2 Simulations and validation

There exist simulation tools for reconstructed trees, where all extinct trees have been pruned off, under homochronous sampling and time dependent rates with the R package TESS (Höhna 2013) and for trees under heterochronous sampling but only with constant rates with the R package TreeSim (Stadler 2011). However, there does not exist a simulation tool for trees under heterochronous sampling and time dependent rates. Since we cannot simulate sampling through time with non-constant rates, we need verify if these models work or not in other ways. We will first examine the likelihood surface for constant rates to see if there seems to be an unique maximum. Secondly, we will simulate trees under constant rates and examine how accurate the inference is. We will choose parameter values so that the simulated trees will be of similar age as the virus phylogenies, in Section 4.1 the data will be presented more thoroughly. For one simulation test we will use the estimated values we got from our analysis on one of the virus trees. By both examine the likelihood and do inference on simulated data we want to be able to say if it is hard to distinguish between the extinction rate μ and the sampling rate ψ . Since μ removes a lineage from the tree and ψ terminate a lineage it could be hard to distinguish between the two. Thirdly, we will simulate trees under constant rates, and then do the inference assuming a periodic speciation rate. We expect the inferred rates to behave in certain ways given that the models work as we want them to, e.g. if we use periodic functions on constant rates we want the period to be really long or the amplitude to be low, to mimic a constant rate.

3.2.1 Likelihood surface

We begin with simulating trees with 400 tips. We put the speciation rate and the extinction rate to known constants, $\lambda = 3$, $\mu = 2.52$ and $\psi = 0.28$. This will often give an age over 10. We chose 400 tips to be a little smaller than the Influenza A phylogenies. We use the R package TreeSim to simulate the trees. First we need to know how the likelihood surface looks, to ascertain that the maximum exist. We put the speciation rate to a fixed (the true) value and vary the extinction rate and sampling rate and calculate the likelihood for each of these values for the simulated tree. Figure 6a and Figure 6b shows the result for one such simulated tree. In this Figure and for others not shown here the results look good, it seems like we have an unique maximum. However, here the true value for λ is fixed and we only need to estimate two parameters.

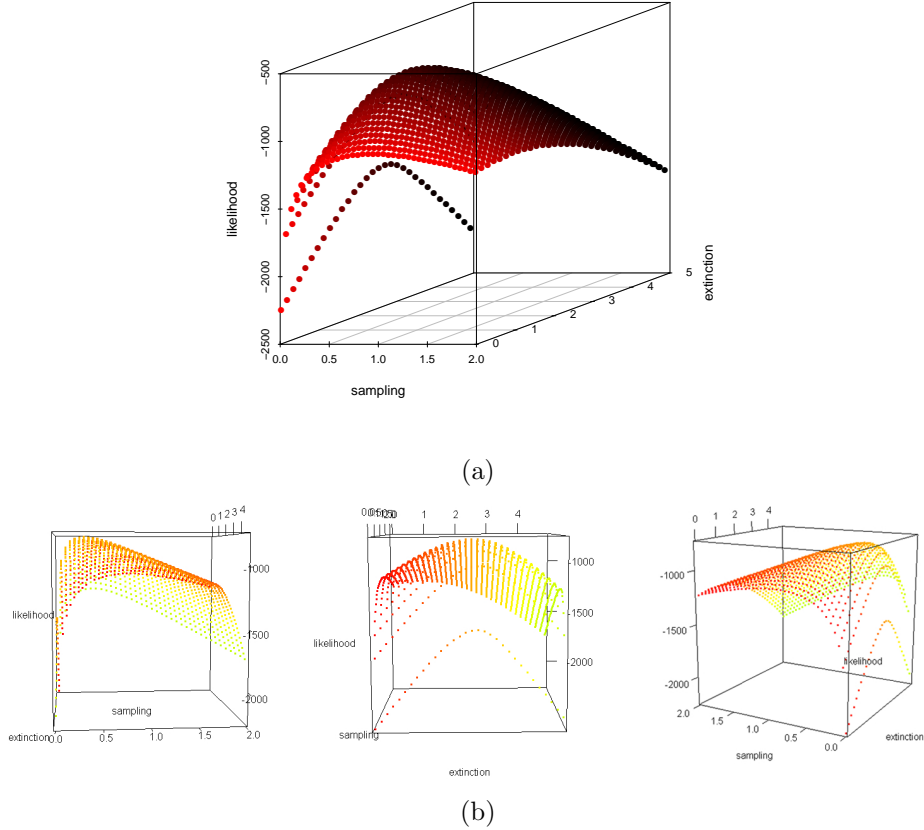


Figure 6: Likelihood surface when λ is fixed and μ and ψ is varied. True values are $\lambda = 3$, $\mu = 2.52$ and $\psi = 0.28$

3.2.2 Simulated data inferred with constant rates

We simulated 500 trees with 600 tips to be of similar size as the virus phylogenies in this thesis. The true rates were $\lambda = 3$, $\mu = 1.89$, $\psi = 0.81$ and $R_0 = 1.11$. The values was chosen because then the age of the trees usually becomes between 8 and 15, a similar age as the virus phylogenies. We analysed these trees assuming constant rates and a summary is shown in table 2. The death rate is the parameter that varies most but R_0 and the total recovery rate $\mu + \psi$ is well estimated. The result is shown in Appendix C and Table C.1 present the results for the 100 first simulations. The relative error for λ is defined as

$$\frac{\sum_{i=1}^{500} \frac{|\hat{\lambda}_i - \lambda|}{\lambda}}{500}.$$

Table 2: Summary simulation when inference was done assuming constant rates

parameter	True value	mean	s.e	rel. error	Q1	median	Q3
λ	3	3.008	0.916	0.255	2.2	2.867	3.589
μ	1.89	1.731	1.179	0.532	0.719	1.621	2.529
ψ	0.81	0.942	0.279	0.301	0.718	0.910	1.166
$\mu + \psi$	2.7	2.673	0.915	0.284	1.874	2.532	3.267
R_0	1.111	1.141	0.064	0.046	1.092	1.129	1.176

We simulated 500 trees with constant rates. All trees have 600 tips.

As already stated the estimates for μ varies some, but the ratio between λ and $\mu + \psi$ is very well estimated. The reason is a correlation between λ and $\mu + \psi$ as can be seen in Figure 7.

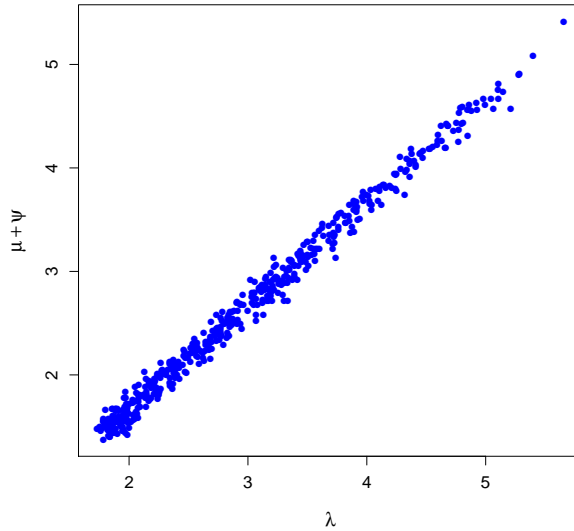


Figure 7: Each point represent the estimated value of λ and the estimated $\mu + \psi$ for the 500 simulated trees when the true values are $\lambda = 3$, $\mu = 1.89$ and $\psi = 0.81$. We see a correlation between estimated λ and $\mu + \psi$.

Additionally to the simulation above we also simulated 500 trees with the same rates as the ones inferred from the phylogeny for the genome segment NA (neuraminidase), assuming constant rates. The number of tips was set to the same as in the phylogenies. The results are shown in Table 4. Here the death rate μ is poorly estimated. There is a correlation between the estimated death rate μ and the estimated sampling rate ψ . If the death rate is small the sampling rate will

compensate by become larger. In Figure 8 the correlation between μ and ψ and also between λ and $\mu + \psi$ is shown. However, the relation between the birth rate and the recovery rate $\mu + \psi$ is well estimated, thus generating good estimates of R_0 .

Table 3: Summary simulation when inference was done assuming constant rates

parameter	True value	mean	s.e	rel. error	Q1	median	Q3
λ	5.6451	4.983	0.861	0.149	4.634	4.816	5.115
μ	2.3590	0.744	1.192	0.767	0.158	0.420	0.899
ψ	3.0703	3.840	0.400	0.264	3.698	3.908	4.077
$\mu + \psi$	5.4293	4.584	0.871	0.185	4.225	4.422	4.698
R_0	1.0397	1.090	0.044	0.051	1.057	1.084	1.117

We simulated 500 trees with constant rates. All trees have 687 tips to be similar in size to the virus phylogenies.

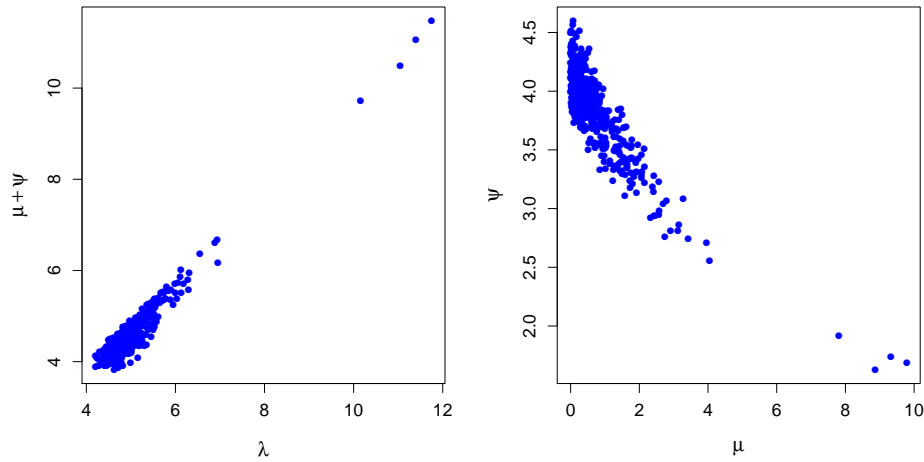


Figure 8: Each point represent an estimated value for the 500 simulated trees when the true values are $\lambda = 5.6451$, $\mu = 2.3590$ and $\psi = 3.0703$. Left figure shows correlation between λ and $\mu + \psi$. Right figure shows correlation between ψ and μ .

3.2.3 Simulated data inferred with periodic speciation rate

We wanted to investigate how often we would chose a periodic speciation rate model in favour of a constant rate model if the underlying data was from constant rates. Therefore we simulated 100 trees and made inference with both the constant rate model and the periodic speciation rate model. The true constant rates was set to $\lambda = 3$, $\mu = 1.89$ and $\psi = 0.81$. A comparison of the AIC values between the models showed that for 88 of the trees a constant rate model would have been chosen. Of the 12 trees where a periodic speciation model is chosen 7 of inferred rates mimics a constant rate and thus leaving us 5 badly inferred cases. The inferred speciation rate for these 12 trees are shown in Figure 9. The likelihood for every of these 100 trees and the two different models are shown in Appendix C and Table C.2.

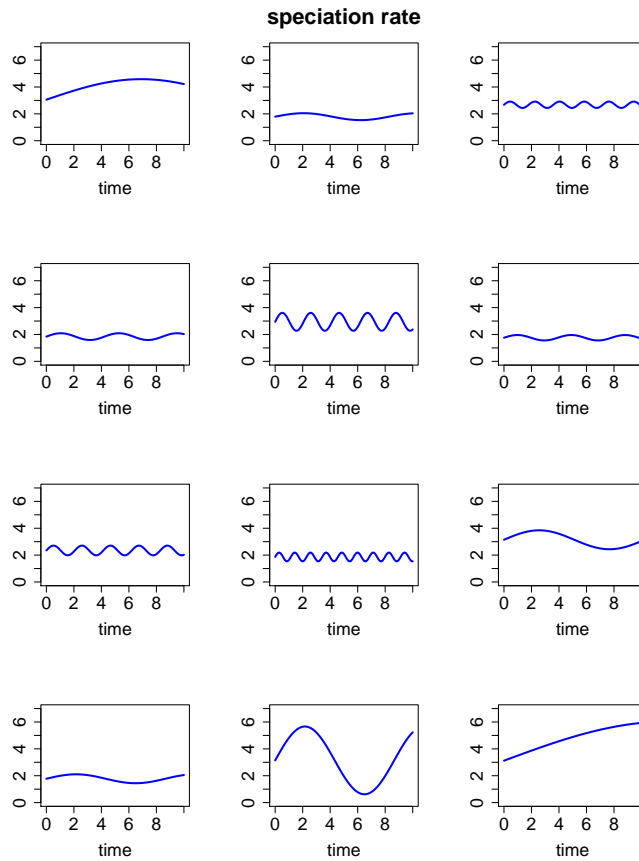


Figure 9: Data was simulated with constant rates. Inference was done assuming both constant rates and then assuming a periodic speciation rate. When comparing AIC values for the two cases 88 trees where best fitted by a constant rate model. For the 12 trees where a periodic speciation rate model was chosen 7 of the inferred speciation rates mimics a constant rate, as shown in this figure.

Then we wanted to determine how the inferred periodic rate behave when the true speciation rate is constant. We simulated additional 100 trees with 600 tips and again the true rates where $\lambda = 3$, $\mu = 1.89$, $\psi = 0.81$. Inference typically gave a very low amplitude for the speciation rate. The mean of the amplitude for the 100 trees is 0.43, the median is 0.197, first quantile 0.089 and the third quantile 0.45. We have a few amplitudes up to a value of 5. In Figure 10 examples of the inferred speciation rate are shown, all amplitudes are plotted in Figure 11. In the cases where the amplitude is large the parameter β , governing the period, is instead small which corresponds to a long period. In Figure 12 we have plotted the amplitude against the period. In all but 6 of the 100 cases we have either a small amplitude or a long period to mimic a constant rate. The inferred R_0 is shown in Figure 13. Sometimes the inferred initial speciation rate is a little below the true value. The initial speciation rate and, thus the value oscillated around has a mean of 2.59, a median of 2.34 and a standard error of 1.01.

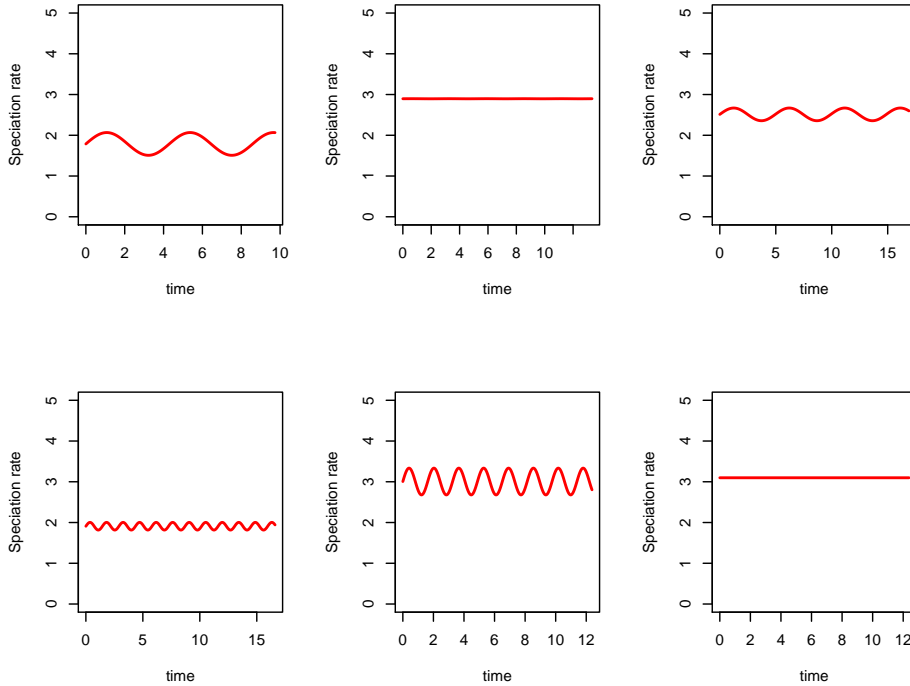


Figure 10: Inferred speciation rate when underlying data is from a constant rate model with $\lambda = 3$. We can see that the inferred speciation rate has a low amplitude

In 11 of the 100 cases the extinction rate μ is estimated to be less than 0.1, but instead the sampling rate is estimated to at least 1.27. In all, it seems like if the phylogenetic tree is generated by constant rates we will discover this both by the

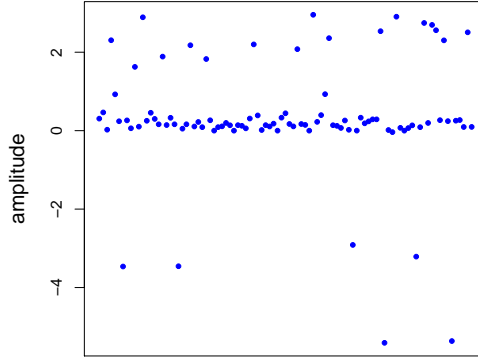


Figure 11: We simulated 100 trees under constant rates and did estimation with a periodic speciation rate. Here the inferred amplitude of the speciation rate is plotted.

amplitude being really low or the period to be very long and otherwise by a model comparison when the constant rate model will be chosen. In table 4 a summary of the parameters able to compare to the true one is shown. The extinction rate μ is not well estimated.

Table 4: Summary of simulation with inference assuming a periodic sciation rate

parameter	True value	mean	s.e	rel. error	Q1	median	Q3
μ	1.89	1.574	1.524	0.685	0.404	1.296	2.041
ψ	0.81	0.976	0.367	0.438	0.7504	0.9907	1.290
$\mu + \psi$	2.7	2.55	1.211	0.356	1.727	2.260	2.777

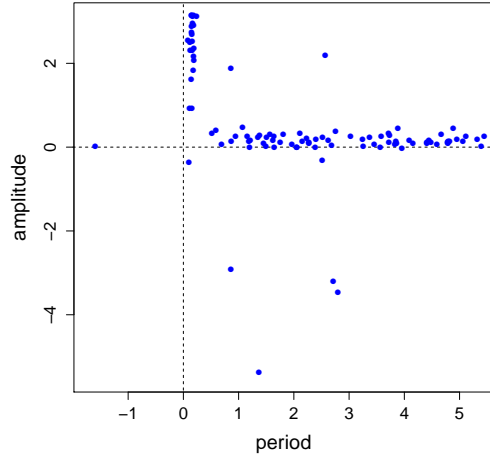


Figure 12: We simulated 100 trees under constant rates and did estimation with a periodic speciation rate. Here the inferred amplitude is plotted against the inferred period. Clearly, if the amplitude α is large the parameter β governing the period becomes very small. A small β gives a long period.

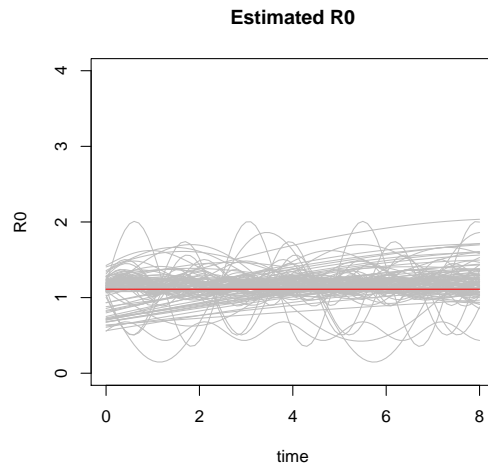


Figure 13: The inferred R_0 for 100 trees simulated under a constant rates but inferred using a periodic speciation rate. The darker line is the true value.

3.2.4 Differential equations

We are using numerical differential equations to obtain the probabilities needed in the likelihood. What we can test is whether these methods obtain a good approximation where an analytical answer exists. In section 2.1.2 we gave the analytical answer for $P_0^S(t, T)$ when $\mu(t) = \psi(t) = 0$ and when $\mu(t) = 0$.

First a comparison of the numerical and analytical answer will be done for the case $\mu(t) = \psi(t) = 0$ and then a comparison for the case $\mu(t) = 0$. We will use $T = 13$ to be in the lower age of the virus phylogenies.

If $\mu(t) = \psi(t) = 0$ we had

$$P_0^S(t, T) = \frac{(1 - \rho)e^{-\int_t^T \lambda(x) dx}}{(1 - \rho)e^{-\int_t^T \lambda(x) dx} + \rho}$$

When $\lambda(t) = \lambda$ the approximation is identical to the analytical values. In Figure 14 two examples are shown. For other values on λ and ρ the result is the same.

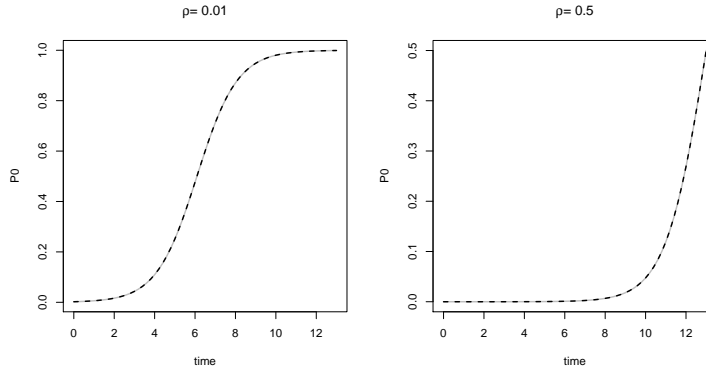


Figure 14: analytical and numerical answer to $P_0^S(t, T)$ in case of $\mu(t) = \psi(t) = 0$ and $\lambda(t) = \lambda$. In these examples $\lambda = 1$. The grey line is the numerical solution and the black line is the analytical solution

However, it is more interesting to test the approximation when we have a periodic speciation rate $\lambda(t) = \alpha \sin(\beta t) + \lambda_m$, $\alpha \leq \lambda_m$. Figure 15 show the result when $\alpha = 0.9$, $\beta = 6$ and $\lambda_m = 1$. We chose $\beta = 6$ since this is close to the value we almost always got from our inference. Here we can see some slight deviations from the true values.

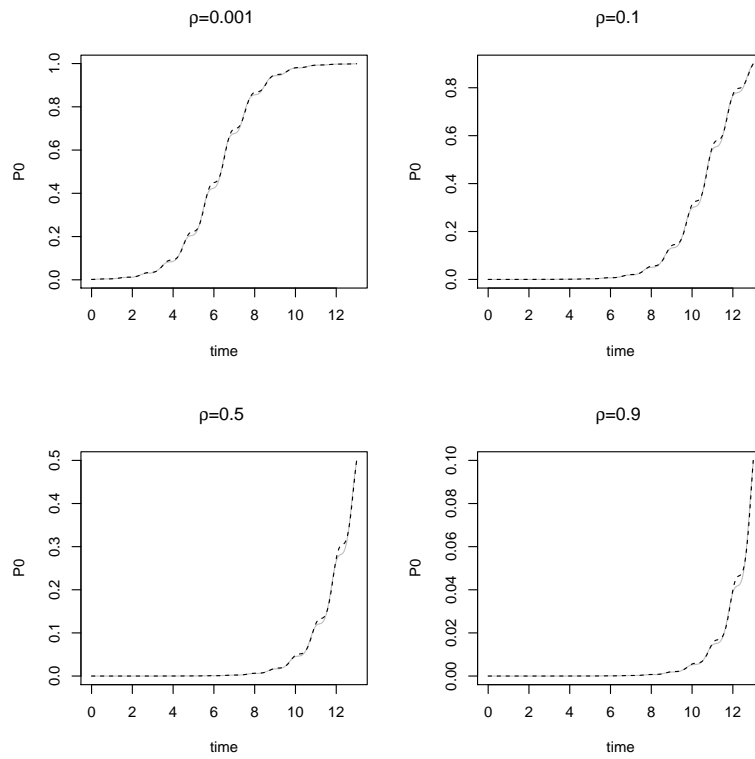


Figure 15: analytical and numerical answer to $P_0^S(t, T)$ in case of $\mu(t) = \psi(t) = 0$ and $\lambda(t) = \alpha \sin(\beta t) + \lambda_m$. In these examples $\alpha = 0.9$, $\beta = 6$ and $\lambda_m = 1$. The grey line is the numerical solution and the black line is the analytical solution

It is for parameter values where the speciation rate is really low together with a very low probability of sampling today that the approximation performs badly. This is the case if ρ is very small and if α is close to λ_m . If we have an α close to λ_m we will often have a speciation rate close to zero. This in combination with a period that yields most time where the speciation rate is close to 0 generates a really bad approximation. That is, parameter values corresponding to a small expected tree size generates a bad approximation. The virus phylogenies we are to make inference on are however not of a small size. If we change λ_m to a higher value we again have a good approximation as shown in Figure 16. However, when we introduce the sampling rate this bad approximation disappears.

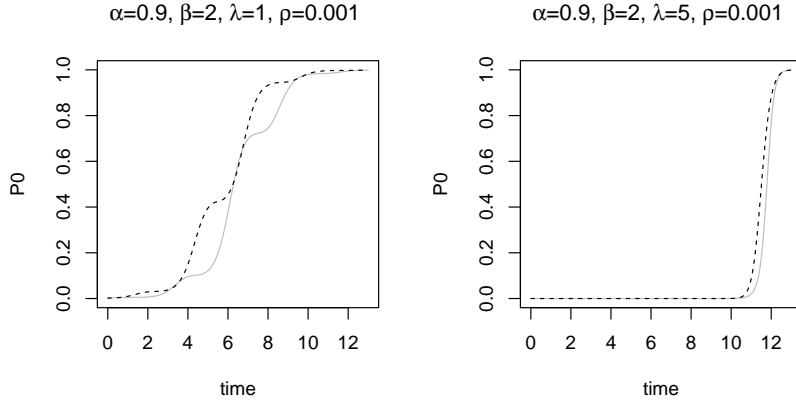


Figure 16: analytical and numerical solution to $P_0^S(t, T)$ in case of $\mu(t) = \psi(t) = 0$ and $\lambda(t) = \alpha \sin(\beta t) + \lambda_m$. The grey line is the numerical solution and the black line is the analytical solution. In the left figure we can see a really bad numerical solution.

Now, let us examine the case when only $\mu(t) = 0$. We had the result

$$P_0^S(t, T) = \frac{(1 - \rho) e^{-\int_t^T \lambda(x) + \psi(x) dx}}{1 - (1 - \rho) \int_t^T e^{-\int_t^x \lambda(s) + \psi(s) ds} \lambda(x) dx}$$

Since we assumed in the analysis $\rho = 0$ we will do the same here. For the previous case this was not relevant since then $P_0^S(t, T)$ was always 1. This seems to work fine as well. In Figure 17 have we plotted some different scenarios where $\alpha = 0.3$ and $\beta = 4$. For more values, see Appendix C and Figure C.4

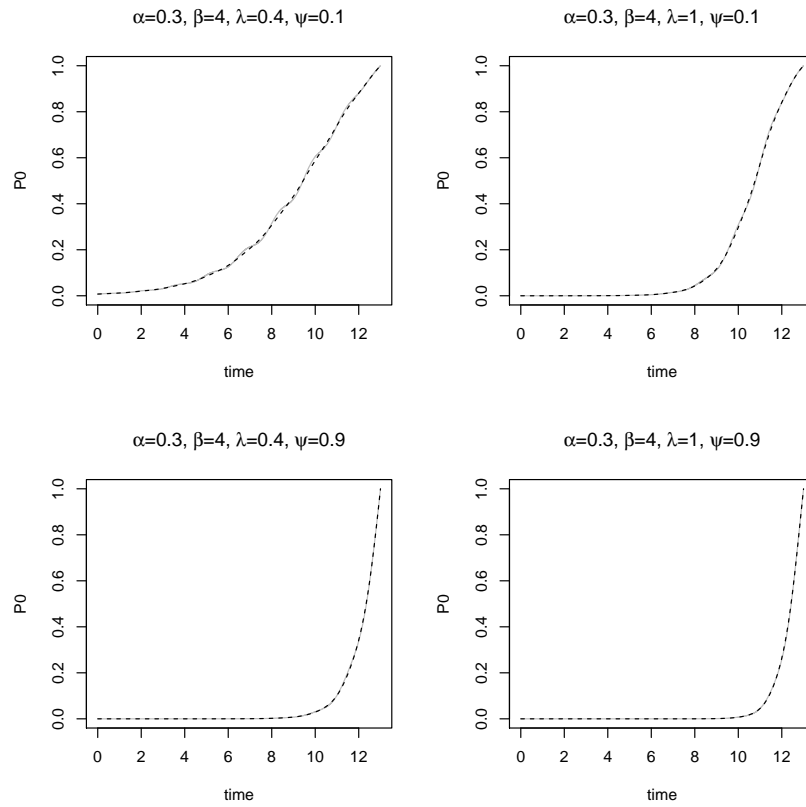


Figure 17: analytical and numerical solution to $P_0^S(t, T)$ in case of $\mu(t) = 0$ and $\lambda(t) = \alpha \sin(\beta t) + \lambda_m$. The grey line is the numerical solution and the black line are the analytical solution

3.3 Inference

The likelihood function, Equation 18, was implemented in R for the given models in Table 1. Optimization and estimation was performed with the function `Optim` and the Nelder-Mead method (Nelder and Mead 1965). The optimization had to be run several times when the periodic rate functions were used. Both numerical integration and differential equation solver was needed and thus even more slowed down the inference. In a maximum likelihood approach the time for optimization is still tractable. R code for the constant rate model and the periodic speciation and extinction model is given in Appendix D. To compare the models we used the Akaike information criterion (AIC)

$$\text{AIC} = 2k - 2\log(L),$$

where k is the number of parameters and L is the maximized likelihood for the model. The chosen model is the one with lowest AIC score.

To estimate the basic reproductive number given constant rates we use the following (Stadler et al. 2012)

$$R_0 = \frac{\lambda}{\mu + \psi}.$$

Additionally, if varying rate functions are used a reproductive number over time can be obtained, $R(t)$, the effective reproductive number. The effective reproductive number is the average number of infections caused by an infectious individual in a population where not everyone is susceptible. R_0 is defined as the average number one initial infective infects when everyone else is susceptible. If this first infected person infects on average more than one other person, if $R_0 \geq 1$, an outbreak is possible. Since R_0 is defined when everyone is susceptible and $R(t)$ is defined when not everyone is necessarily susceptible we have that $R(t) \leq R_0$ and if a large proportion of the population is susceptible in the beginning of an infection $\max R(t) \approx R_0$.

4 Analysis

4.1 Data

The influenza virus is a single-stranded (negative) RNA virus of the family *Orthomyxoviridae*. One of the most important characteristic of RNA viruses is their very high mutation rate (Drake et al. 1998). In this study we will use sequence-data from Influenza A virus, which causes the majority of influenza infections. Different types of influenza A virus can infect humans, birds, poultry and other mammals. Some well known variants are viruses causing bird flu and swine flu. Influenza A virus are categorised by their membrane surface glycoproteins hemagglutinin (HA) and neuraminidase (NA), e.g. H5N1 bird flu. Hemagglutinin is responsible for the binding to host cells and neuraminidase helps in the virion release from host cells. The Influenza genome comprises eight segments. The data we have are from these eighth segments, two from HA and NA proteins, additionally we have data for M1 (matrix protein, envelope of the virus) and M2 protein (ion channel protein) labelled M1/2, three different viral polymerase PA, PB1 and PB2 (associated with replication and transcription), nucleocapsid protein NP and non-structural proteins NS1 and NS2 labelled NS1/2.

The models explained in Section 3.1 were applied to the 8 different phylogenetic trees of the genome segments of H3N2 human influenza A virus from New York State (USA) from Rambaut et al. (2008). The different segment trees used in this thesis is summarized in Table 5 and the phylogenies are plotted in Appendix B.

Table 5: The 8 different genome segments. All phylogenies have 687 tips

Name	Age
1. NA	14.684
2. HA	12.961
3. PA	13.719
4. PB1	15.583
5. PB2	14.022
6. M1/2	12.509
7. NP	13.833
8. NS1/2	12.657

4.2 Results

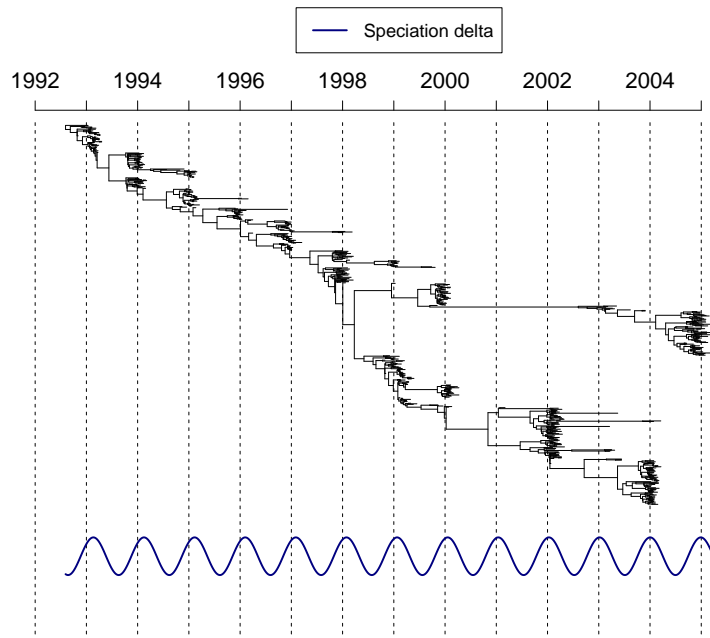
The result from the AIC comparison is shown in Table 6. One main result is that if a periodic rate function is included for the speciation or extinction rate we get a higher likelihood. The likelihood is often increased so much it cannot only be explained by the additional parameters.

Table 6: Chosen models and the estimated parameter values for each phylogeny.

Name	Periodic Model	λ_m/λ	μ_m/μ	ψ	α_1	β_1	α_2	β_2	δ
NA	speciation	9.50	6.22	1.39	-8.49	6.499			
HA	speciation shift	6.699	4.57	1.28	-5.976	6.33			0.409
PA	speciation shift	9.33	6.958	1.45	7.88	6.53			3
PB1	speciation shift	7.27	4.06	2.2			-4.058	6.55	1.04
PB2	spec. and ext.	5.23	5.26	2.37	-4.27	6.32	5.08	6.16	
M1/2	extinction shift	18.58	12.69	2.32			8.89	6.04	2.0
NP	spec. and ext.	9.94	10.04	3.10	-9.31	2.28	9.11	6.28	
Ns1/2	speciation shift	7.074	4.29	2.55	-5.68	6.36			1.26

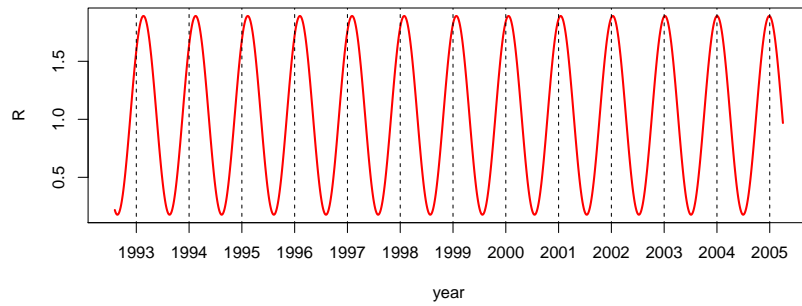
All phylogenies is best explained by a periodic rate model. The period is also almost every time estimated close to a yearly period.

The parameter β representing the period is always estimated to a value close to 6.3 which is the value corresponding to yearly period. As an example we will show the periodic rates for the best fitting model for the Ns1/2 and NP phylogeny in Figure 18 and Figure 19. The best fitting model for the Ns1/2 phylogeny is the one with a periodic speciation rate and a constant extinction rate and the best fitting model for the NP phylogeny is the one with both periodic speciation and extinction rate. The inferred period for Ns1/2 is almost exactly yearly. We can see that every big new speciation period is correctly inferred. We can also see that the periodic extinction rate fits well in the NP phylogeny. All phylogenies and the inferred rates for the best fitted model can be found in Appendix B. At first one can wonder why the speciation rate does not have a yearly period for the NP phylogeny, but then examine the estimated $R(t)$ in Figure 19. In contrast to a periodic speciation rate and a constant extinction rate where the same values return periodically we can here have a $R(t)$ that varies between years. For example, in year 2003 there are very few new speciation events, and this is captured in the estimated $R(t)$.



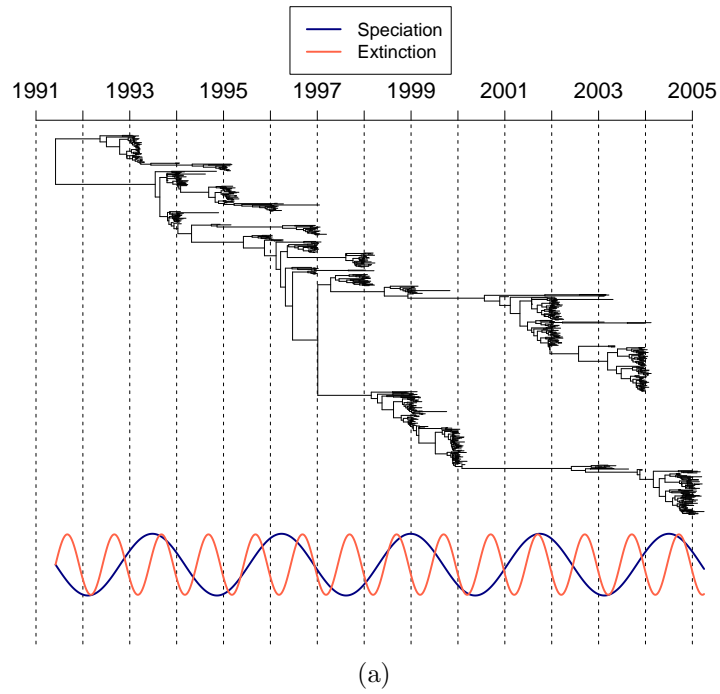
(a)

Ns12 phylogeny



(b)

Figure 18: The best fitted model for the Ns1/2 phylogeny was the periodic speciation model. (a) The phylogeny and the best fitted periodic rate. (b) Estimated $R(t)$ from the best fitted model



NP phylogeny

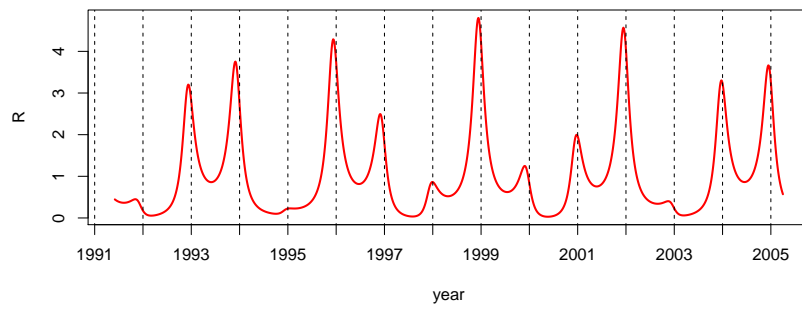


Figure 19: The best fitted model for the NP phylogeny was the periodic speciation and extinction model. (a) The phylogeny and the best fitted periodic rate. (b) Estimated $R(t)$ from the best fitted model

Table 7: Summary of estimated $R(t)$ for the best fitted models

Phylogeny	NA	HA	PA	PB1	PB2	M1/2	NP	Ns1/2
Min	0.1328	0.1236	0.1725	0.7046	0.0752	0.7774	0.0295	0.1775
Max	2.3642	2.1667	2.04680	3.3015	3.7056	3.0359	4.7629	1.8909

All segments except PB2 and NP have the same maximum of $R(t)$ every year. The maximum of $R(t)$ for PB2 ranges from 2.4565 to 3.7055. For the NP segment the maximum of $R(t)$ each year varies from values below 1 year 1995, 1998 and 2003 up to the highest value of 4.7629 year 1999.

5 Discussion

The two goals of this thesis were to investigate if we can detect the seasonal behaviour of influenza virus from its sequence data and to estimate the effective reproductive number. This periodical behaviour can be detected and our estimated effective reproductive number is in line with previous studies. In the next section we will discuss this further and we will also discuss the difficulties using our models.

5.1 Reproductive number

We have inferred the ratio between the birth rate $\lambda(t)$ and the recovery rate $\mu(t) + \psi$ for the eight different segments of H3N2 human influenza A virus. Thus we have eight different reproductive rates for the same virus. To be able to do a comparison with the results of other studies, we calculated the mean of all yearly maximal $R(t)$ and they are shown in Table 8. It is not obvious that we can take the mean of all the segments maximal $R(t)$ to get a good representing value. However, the effective reproductive number $R(t)$ for the different segments do not differ dramatically from each other except for the two segments PB2 and NP. The estimated $R(t)$ of segment PB2 is similar to the others in early years but then increases. For the segment NP the mean of the yearly maximal estimated $R(t)$ is 2.647.

Table 8: Mean of maximum of $R(t)$ each year

1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
2.582	2.676	2.264	2.789	2.585	2.398	2.898	2.466	2.562	2.881	2.357	2.710	2.743

The estimates in Table 8 are as high or higher than the values of the effective reproductive number obtained for seasonal influenza according to Chowell et al. (2008) who reported values of 0.9-2.1. However, many different values are reported

and Dushoff et al. (2004) claim that appropriate values of R_0 for the seasonal influenza are 4-16, while Keeling and Rohani (2008) report R_0 of 3-4. Barnea et al. (2011) who have done analyses on H3N2 subtype, but in Israel, get $R_0 = 2.7 - 8.0$ but a lower value for the effective reproductive number of 1.12-1.33. They define $R_{eff} = R_0 S_0 / N$ where S_0 is the initial number of susceptibles in the population and N the population size. In all, this indicates that our estimates here are quite reasonable.

5.2 Yearly Period

In this study, the inferred period for influenza virus is close to a yearly one. This yearly period could be an artefact of the sampling strategy of the influenza virus. In New York State, USA, the influenza season is during winter time. It is within this time span the sampled virus data come from. If someone gets sick with influenza outside this period he or she is unlikely to be diagnosed with influenza and in this case would not come into our sample. We will need better simulation tools to verify if this could be the reason for the yearly period.

What we did was additional analyses on a phylogenetic tree of the measles virus reconstructed from samples between 1979 and 2009 (Bedford et al. 2011). There was exactly one year in between samplings between the years 1996 and 2009. We were interested if we for the measles phylogeny would get an inferred yearly period as well. To do inference, a model with a periodic speciation rate and a constant extinction rate was used. The inferred period for this tree was 3.3 years. This indicates that sampling may not influence the inferred period and that it is possible to detect the seasonal behaviour of viruses from its sequence data.

5.3 Studied models and possible drawbacks

We have used different periodic functions for the speciation and extinction rate. The reason for this was that we wanted to model the periodic outbreaks of influenza virus. More specifically the sine function was used for the speciation rate and the cosine function for the extinction rate. One restraint of the periodic functions used in this thesis is that every periods maximum and minimum will be the same. It is, however, very likely that the maximum of the speciation rate for different years will differ in reality.

When trying to infer the periodic rates we need to run the optimization several times to get the global maximum likelihood. When using the sine function and the given parameterization, several different values of the parameters can give the same actual rate, since $-\alpha \sin(-\beta t) = \alpha \sin(\beta t)$. That is not the only difficulty. For example, let us examine the periodic speciation model without a shift. If we fix all parameter values (MLE's) except the parameter β governing the period and

examine the likelihood, we typically get several local optima as in Figure 20. Even though the value corresponding to a yearly period gives the highest likelihood we have several local maxima where the optimization can get stuck.

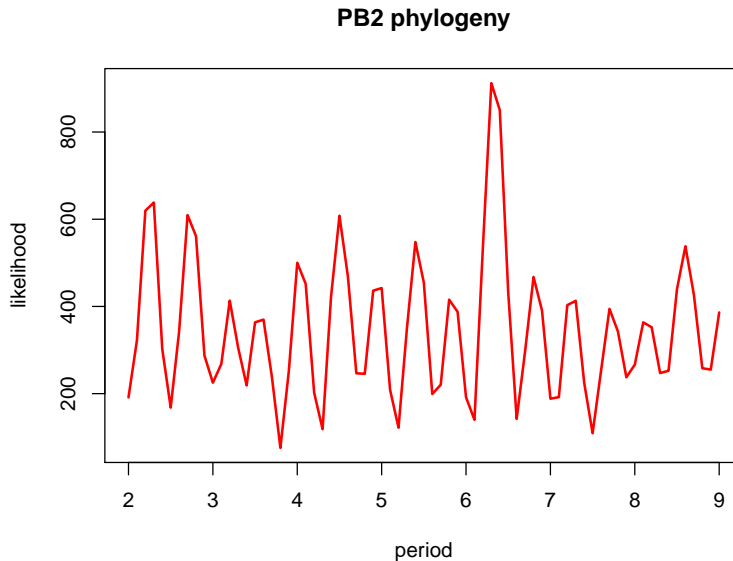


Figure 20: Likelihood for PB2. We have fixed every parameter except β governing the period and plotted the likelihood for different values of β

As said in Section 3.2.2 we cannot verify that the estimates will be correctly estimated if the underlying rates are periodic, since we cannot do simulations. However, for constant rates the likelihood function produces reliable estimates, especially of R_0 . If the inference is done assuming periodic rates we get estimates relatively close to the true values. In most of the 100 trees the inferred speciation rate either has a low amplitude or a long period to mimic a constant rate. The constant rate model is also chosen in favour of the periodic speciation model if the data has been simulated under constant rates.

One problem previously addressed in section 3.2.4 is the numerical solutions to the probability $P_0^S(t, T)$. The approximation worked fine, except for values when we had a prolonged and very low speciation rate, a very low probability of being sampled today and no sampling rate. Even though this is concerning, the most important thing for us here is that the approximation works fine when the sampling rate was introduced.

6 Summary and future work

6.1 Summary and conclusions

In this thesis we have used birth and death processes on phylogenetic trees reconstructed from heterochronous sequencing. More importantly, we have tried to extend this to time dependent rates, specifically periodic ones. We derived the likelihood and used this to perform inference on influenza A virus phylogenies for eight different gene segments from New York State, USA. The influenza virus is known to come in seasons and we were interested if we could detect this from its sequence data. As a main result, we always chose a periodic rate function over the constant rate function. The inferred period is always close to a yearly one. This could be an artefact of the sampling strategy but we need better simulation tools to verify if this is the case. We are able to estimate the effective reproductive number over time $R(t)$. The mean of the maximum estimated effective reproductive number of these 8 segments is at most 2.88. The inferred effective reproductive number is also in line with previous studies, even if it is hard to compare the results when we do not use the exact same datasets. Furthermore, the data here is a set of sampled viruses which leads to that the speciation/infection rate is expected to be lower than inferred infection rates in studies using additional data.

One important aspect of being able to use time dependent rates is that we can allow for periods of higher extinction rate than speciation rate. This is important not only to be able to model the initial exponential increase of an epidemic, as is the case for constant rates, but also the process when the number of susceptibles in the population decreases. This is something we are able to catch with our periodic time dependent rates.

6.2 Future work

The most important work to be done is to create ways of simulating trees with time dependent rates and heterochronous sampling to be able to verify how well our models perform. If we assume that sampling means a termination of a branch and that the extinction rate and sampling rate is constant, this should not be too difficult. To do this, we need to be able to simulate complete trees with a time dependent speciation rate. Then, at every extinction event we remove the lineage with a probability $\frac{\mu}{\mu+\psi}$ and keep the extinct lineage with a probability $\frac{\psi}{\mu+\psi}$, the latter becoming a sampling event.

In this thesis we assume that the probability of sampling today, ρ , is 0. Instead of this, it is possible to do analysis assuming different values on ρ . We started this for the periodic speciation rate model, but due to time limitations it was not possible to do it for every model and with enough different values on ρ .

Another assumption made in this thesis is that a sampling is equivalent to a removal of the lineage. This could be a too big simplification of the reality for influenza virus. But on the other hand, we usually do not get infected by the exact same virus strain twice. Moreover, this assumption makes it possible for us to simulate trees with constant rates and heterochronous sampling. We believe that multiple sampling is most relevant for fossils rather than influenza virus. Despite all this, we still believe one should try to use multiple sampling for different virus phylogenies. Multiple sampling makes sense for viruses which stays a long time in the host, such as HIV and HSV (herpes simplex virus).

It would be interesting to do inference with time dependent rates on other types of diseases, where something other than a periodic rate function would make sense. For example, for HIV we could expect the saturation phase just to be reached which makes it different from the influenza virus. We could also expect a big fraction of individuals infected with HIV being sampled, further distinguishing it from the influenza virus studied in this thesis.

References

- O. Barnea, R Yaari, G. Katriel, and L Stone. Modelling seasonal influenza in israel. *Mathematical biosciences and engineering.*, 8(2):561–573, 2011.
- T Bedford, S Cobey, and M Pascual. Strength and tempo of selection revealed in viral gene genealogies. *BMC Evolutionary Biology*, 11(1):220, 2011. ISSN 1471-2148.
- G. Chowell, M. A. Miller, and C. Viboud. Seasonal influenza in the united states, france, and australia: transmission and prospects for control. *Epidemiology & Infection*, 136:852–864, 5 2008. ISSN 1469-4409.
- G. Didier, M. Royer-Carenzi, and M. Laurin. The reconstructed evolutionary process with the fossil record. *Journal of Theoretical Biology*, 315:26–37, 2012.
- J.W Drake, B Charlesworth, D Charlesworth, and J.F Crow. Rates of spontaneous mutation. *Genetics Society of America*, 148:1667–1686, 1998.
- Jonathan Dushoff, Joshua B. Plotkin, Simon A. Levin, and David J. D. Earn. Dynamical resonance can account for seasonality of influenza epidemics. *Proceedings of the National Academy of Sciences of the United States of America*, 101(48):16915–16916, 2004.
- A. W. F. Edwards. Estimation of the branch points of a branching diffusion process. *Journal of the Royal Statistical Society. Series B*, 32(2):155–174, 1970.
- D. Hansson. Is diversification limited? a birth-death model comparison. 2011.
- S. Höhna. Fast simulation of reconstructed phylogenies under global, time-dependent birth-death processes. *Bioinformatics*, 2013.
- M.J. Keeling and P Rohani. *Modeling Infectious Diseases in humans and animals*. Princeton University Press, 2008.
- DG. Kendall. Stochastic processes and population growth. *Journal of the Royal Statistical Society. Series B (Methodological)*, 11(2):230–264, 1949.
- G.E. Leventhal, R. Kouyos, T. Stadler, V. von Wyl, S. Yerly, C. Celleraï, T. Klimkait, H.F. Gnthard, and S. Bonhoeffer. Inferring epidemic contact structure from phylogenetic trees. *Philosophical Transactions: Biological Sciences*, 8(3), 2012.
- H. Morlon, T. Parsons, and T. Plotkin. Reconciling molecular phylogenies with the fossil record. *Proceedings of the National Academy of Sciences*, 108(39):16327–16332, 2011.

- S. Nee, M.A. May, and P.H. Harvey. The reconstructed evolutionary process. *Philosophical Transactions: Biological Sciences*, 344(1309):305–311, 1994.
- J. A. Nelder and R. Mead. A simplex algorithm for function minimization. *Computer Journal*, 7:308–313, 1965.
- L. Orlando, D. Bonjean, H. Bocherens, A. Thenot, A. Argant, M. Otte, and C Hnni. Ancient dna and the population genetics of cave bears (*ursus spelaeus*) through space and time. *Mol. Biol. Evol*, 2002.
- O.G. Pybus, M.A. Charleston, S. Gupta, A. Rambaut, E.C. Holmes, and P.H. Harvey. The epidemic behavior of the hepatitis c virus. *Science*, 2001.
- A. Rambaut, O.G. Pybus, M.I. Nelson, J.K. Taubenberger, and C.H. Holmes. The genomic and epidemiological dynamics of human influenza a virus. *Nature*, 453: 615–620, 2008.
- Fredrik Ronquist, Seraina Klopstein, Lars Vilhelmsen, Susanne Schulmeister, Debra L. Murray, and Alexandr P. Rasnitsyn. A total-evidence approach to dating with fossils, applied to the early radiation of the hymenoptera. *Systematic Biology*, 2012.
- T Stadler. Sampling-through-time in birth-death trees. *Journal of Theoretical Biology*, 267:396–404, 2010.
- T. Stadler. Simulating trees with a fixed number of extant species. *Systematic Biology*, 2011.
- T. Stadler, R. Kouyos, V. von Wyl, S. Yerly, J. Bni, P. Brgisser, T. Klimkait, B. Joos, P. Rieder, D. Xie, H.F. Gnthard, A.J. Drummond, S. Bonhoeffer, and Swiss HIV Cohort Study. Estimating the basic reproductive number from viral sequence data. *Mol. Biol. Evol*, 29:347–357, 2012.
- T. Stadler, D. Khner, S. Bonhoeffer, and J.A. Drummond. Birthdeath skyline plot reveals temporal changes of epidemic spread in hiv and hepatitis c virus (hcv). *Proceedings of the National Academy of Sciences*, 110(1):228–233, 2013.
- EA. Thompson. *Human evolutionary trees*. Cambridge University Press, Cambridge, 1975.
- S.S. Twiddy, E.C. Holmes, and A Rambaut. Inferring the rate and time-scale of dengue virus evolution. *Mol. Biol. Evol*, 2003.
- Erik M. Volz, Sergei L. Kosakovsky Pond, Melissa J. Ward, Andrew J. Leigh Brown, and Simon D. W. Frost. Phylodynamics of infectious disease epidemics. *Genetics*, 183(4):1421–1430, 2009.

A Equality between Equation 2 and Thompson (1975)

In Section 2.1.1 we claim that the probability of having N species at a time T if we had one species at time t , $P_N(t, T)$, in Equation 2 becomes the same as in Thompson (1975) if we have constant rates $\lambda(t) = \lambda$, $\mu(t) = \mu$ and $\rho = 1$. The equations where

$$\begin{aligned} P_0(0, t) &= 1 - P_S(0, t) \\ P_1(0, t) &= P_S(0, t)^2 e^{r(0, t)} \\ P_N(0, t) &= P_S(0, t)^2 e^{r(0, t)} \left(1 - P_S(0, t) e^{r(0, t)}\right)^{N-1} \end{aligned}$$

First, when considering constant rates it is the length of time that is important since $P_N(t + x, T + x)$ is the same for all values of x . We will therefor instead use the notation $P_N(s)$ where $s = T - t$. The probabilities of N descendants from a single ancestor after a time s in Thompson (1975) are

$$\begin{aligned} p_0(s) &= \frac{\mu(1 - e^{-(\lambda-\mu)s})}{\lambda - \mu e^{-(\lambda-\mu)s}} \\ p_1(s) &= \frac{(\lambda - \mu)^2 e^{-(\lambda-\mu)s}}{(\lambda - \mu e^{-(\lambda-\mu)s})^2} \\ p_N(s) &= (\lambda/\mu)^{N-1} p_1(s) p_0(s)^{N-1} \end{aligned}$$

Our Equation (3) with constant rates becomes

$$r(t, T) = -(\lambda - \mu)(T - t) = -(\lambda - \mu)s,$$

and Equation (4) becomes

$$\begin{aligned} P_S(t, T) &= \left(1 + \int_t^T \mu e^{r(t, \tau)} d\tau\right)^{-1} = \left(1 + \int_t^T \mu e^{-(\lambda-\mu)(\tau-t)} d\tau\right)^{-1} \\ &= \frac{1}{1 + \frac{-\mu e^{-(\lambda-\mu)(T-t)} + \mu}{\lambda - \mu}} = \frac{\lambda - \mu}{\lambda - \mu e^{-(\lambda-\mu)(T-t)}} = \frac{\lambda - \mu}{\lambda - \mu e^{-(\lambda-\mu)s}}. \end{aligned}$$

Now it is straightforward to show that the probabilities are the same for constant rates.

$$\begin{aligned} P_0(t, T) &= 1 - P_S(t, T) = 1 - \frac{\lambda - \mu}{\lambda - \mu e^{-(\lambda-\mu)s}} = \frac{\mu - \mu e^{-(\lambda-\mu)s}}{\lambda - \mu e^{-(\lambda-\mu)s}} = p_0(s) \\ P_1(t, T) &= P_S(t, T)^2 e^{r(0, t)} = \left(\frac{\lambda - \mu}{\lambda - \mu e^{-(\lambda-\mu)s}}\right)^2 e^{-(\lambda-\mu)s} = p_1(s) \end{aligned}$$

And finally we show that $P_N(t, T) = P_N(s) = p_N(s)$.

$$\begin{aligned}
P_N(t, T) &= P_S(t, T)^2 e^{r(t, T)} \left(1 - P_S(t, T) e^{r(t, T)}\right)^{N-1} = P_1(t, T) \left(1 - P_S(t, T) e^{r(t, T)}\right)^{N-1} \\
&= p_1(s) \left(1 - \frac{(\lambda - \mu) e^{-(\lambda - \mu)s}}{\lambda - \mu e^{-(\lambda - \mu)s}}\right)^{N-1} = p_1(s) \left(\frac{\lambda - \lambda e^{-(\lambda - \mu)s}}{\lambda - \mu e^{-(\lambda - \mu)s}}\right)^{N-1} \\
&= p_1(s) \left(\frac{\lambda}{\mu} \cdot \frac{\mu - \mu e^{-(\lambda - \mu)s}}{\lambda - \mu e^{-(\lambda - \mu)s}}\right)^{N-1} = p_1(s) \left(\frac{\lambda}{\mu} \cdot p_0(s)\right)^{N-1} = p_N(s).
\end{aligned}$$

B Complete Results

Table B.1: Results constant speciation and no extinction

Name	LogL	$\hat{\lambda}$	$\hat{\psi}$	\hat{R}_0
NA	626.25	4.27	4.06	1.0533
HA	624.31	4.27	4.06	1.0498
PA	650.44	4.35	4.14	1.0504
PB1	519.58	3.96	3.75	1.0544
PB2	462.04	3.80	3.60	1.0563
M1/2	1086.14	5.96	5.79	1.0301
NP	678.29	4.44	4.22	1.0507
Ns1/2	1034.31	5.74	5.594	1.0259

Table B.2: Results constant rates

Name	LogL	$\hat{\lambda}$	$\hat{\mu}$	$\hat{\psi}$	\hat{R}_0
NA	626.25	4.48	0.40	3.87	1.0507
HA	624.31	4.47	0.38	3.88	1.0475
PA	650.44	4.40	0.11	4.09	1.0498
PB1	519.58	3.98	0.04	3.73	1.0541
PB2	462.04	3.91	0.22	3.49	1.0546
M1/2	1086.14	7.55	2.81	4.57	1.0236
NP	678.29	4.59	0.30	4.08	1.0490
Ns1/2	1034.31	5.79	0.10	5.54	1.0257

Table B.3: Results periodic speciation and no extinction

Name	LogL	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\lambda}_m$	$\hat{\psi}$
NA	660.65	-2.07	6.57	3.79	4.03
HA	690.30	-2.55	6.47	3.55	4.05
PA	687.90	1.93	6.28	3.89	4.13
PB1	558.49	1.97	6.36	3.76	3.70
PB2	523.77	-2.27	6.41	3.28	3.54
M1/2	1108.16	1.77	6.42	5.22	5.78
NP	737.94	-2.55	6.49	3.66	4.22
Ns1/2	1063.67	2.68	6.28	4.90	5.57

Table B.4: Results periodic birth rate

Name	LogL	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\lambda}_m$	$\hat{\mu}$	$\hat{\psi}$
NA	1230.95	-8.49	6.50	9.50	6.22	1.39
HA	1052.02	-4.80	6.38	6.02	4.33	2.34
PA	973.02	-5.44	6.50	6.56	4.14	1.74
PB1	681.07	-3.75	6.53	6.22	4.11	2.00
PB2	947.39	-4.77	6.34	6.60	5.02	1.73
M1/2	1110.84	-1.90	-6.42	5.75	1.13	5.27
NP	1046.15	-4.18	6.42	7.92	6.11	1.71
Ns1/2	1121.27	-3.65	6.53	7.50	3.92	3.80

Table B.5: Results periodic birth rate with shift δ

Name	LogL	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\lambda}_m$	$\hat{\delta}$	$\hat{\mu}$	$\hat{\psi}$
NA	1219.83	7.75	6.47	10.30	3.14	7.16	1.33
HA	1214.56	-5.98	6.33	6.70	0.41	4.57	1.28
PA	1212.98	7.88	6.53	9.34	3.00	6.96	1.45
PB1	665.53	-3.19	6.45	6.87	0.81	5.26	1.67
PB2	787.62	-5.32	6.36	5.05	-0.11	2.34	1.24
M1/2	1111.15	-2.30	6.44	5.73	2.97	1.11	5.38
NP	1255.84	7.60	6.24	9.23	-1.67	8.55	1.60
Ns1/2	1243.05	-5.68	6.36	7.07	1.26	4.30	2.55

Table B.6: Results periodic death rate

Name	LogL	$\hat{\lambda}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\mu}_m$	$\hat{\psi}$
NA	1078.74	16.40	13.65	6.32	13.91	0.87
HA	1175.87	8.19	7.0981	6.17	7.0986	2.39
PA	834.46	9.65	-4.30	6.60	7.69	2.10
PB1	683.93	7.64	4.91	6.37	4.92	1.95
PB2	825.00	6.60	3.8908	6.15	3.8915	2.14
M1/2	1242.45	10.91	-5.11	6.25	6.96	3.37
NP	1184.40	10.92	11.05	6.24	11.05	2.97
Ns1/2	1200.43	7.11	2.97	6.29	2.97	4.42

Table B.7: Results periodic death rate with shift δ

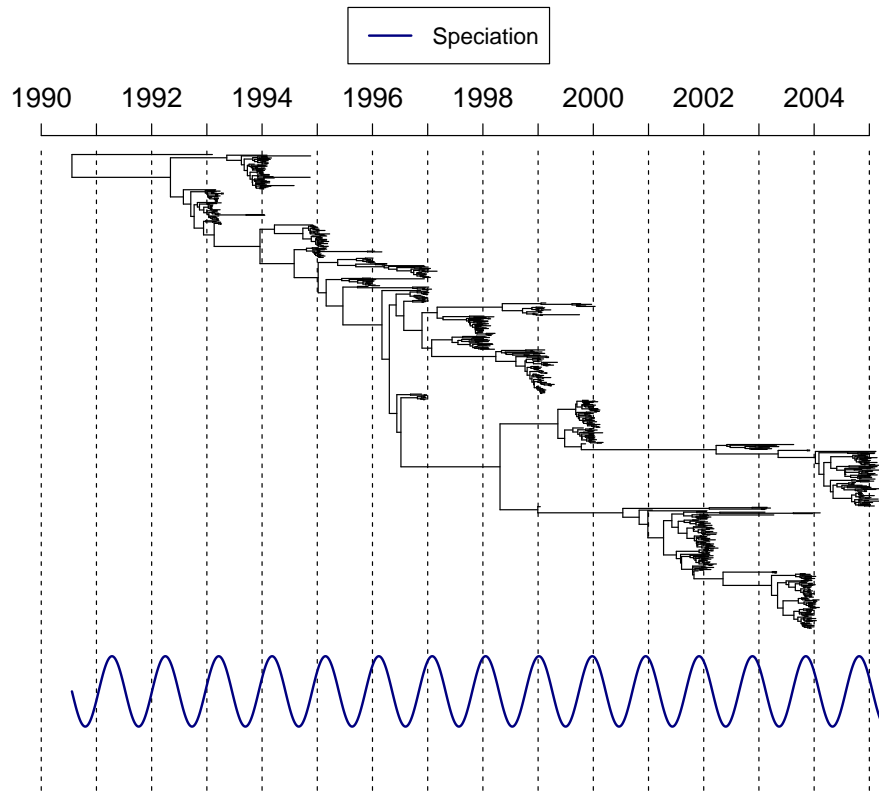
Name	LogL	$\hat{\lambda}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\mu}_m$	$\hat{\delta}$	$\hat{\psi}$
NA	814.44	11.89	-7.80	7.03	7.81	0.66	1.63
HA	884.72	5.13	2.006	6.28	2.009	-0.97	2.90
PA	859.86	9.71	-4.31	6.60	6.71	0.08	2.00
PB1	734.08	7.27	-4.0576	6.55	4.0577	1.04	2.20
PB2	978.42	6.80	3.976	6.32	3.978	-1.61	2.05
M1/2	1320.88	18.58	-8.89	6.04	12.70	2.00	2.32
NP	1008.26	8.10	3.67	-6.37	4.30	1.35	2.59
Ns1/2	1160.73	6.72	2.359	6.46	2.369	-1.50	3.98

Table B.8: Results periodic birth rate and death rate

Name	LogL	$\hat{\alpha}_1$	$\hat{\beta}_1$	$\hat{\lambda}_m$	$\hat{\alpha}_2$	$\hat{\beta}_2$	$\hat{\mu}_m$	$\hat{\psi}$
NA	685.26	2.16	6.26	4.37	-1.50	2.71	1.51	3.43
HA	820.35	12.54	5.92	13.48	5.17	0.79	11.70	1.38
PA	959.64	0.48	2.10	6.28	4.52	6.31	4.53	2.21
PB1	679.00	-3.67	4.02	7.16	5.48	5.12	5.48	1.97
PB2	1281.74	-4.27	6.32	5.22	5.08	6.16	5.26	2.37
M1/2	1107.91	-1.30	1.44	6.58	0.70	2.37	0.70	5.31
NP	1405.86	-9.31	2.28	9.94	9.11	6.26	10.04	3.10
Ns1/2	1062.85	2.71	6.29	4.9	-0.0003	6.62	0.0004	5.79

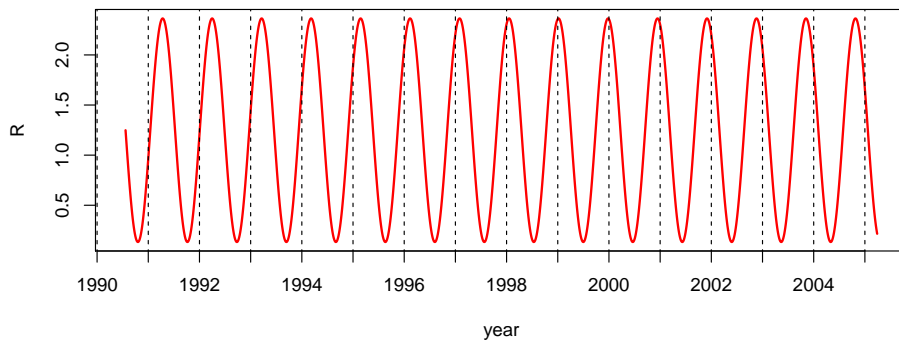
Table B.9: AIC

Name	CS	CS & CE	PS & CE	PS	PS δ & CE	CS & PE	CS & PE δ	PS & PE
NA	-1248.5	-1246.5	-1313.3	-2451.9	-2427.7	-2147.5	-1616.8	-1356.52
HA	-1244.6	-1242.7	-1372.6	-2094.0	-2417.4	-2341.8	-1757.4	-1626.7
PA	-1296.9	-1294.9	-1367.8	-1936.04	-2413.9	-1659.0	-1707.7	-1905.28
PB1	-1035.2	-1033.2	-1109.0	-1352.1	-1319.1	-1357.8	-1456.2	-1344
PB2	-920.1	-918.1	-1039.5	-1884.8	-1563.2	-1640.0	-1944.8	-2549.48
M1/2	-2168.3	-2166.4	-2208.3	-2211.7	-2210.3	-2474.8	-2629.8	-2201.82
NP	-1352.6	-1350.6	-1467.9	-2082.3	-2499.7	-2358.8	-2004.5	-2797.72
Ns1/2	-2064.6	-2062.7	-2119.1	-2232.5	-2474.1	-2390.8	-2309.5	-2111.7



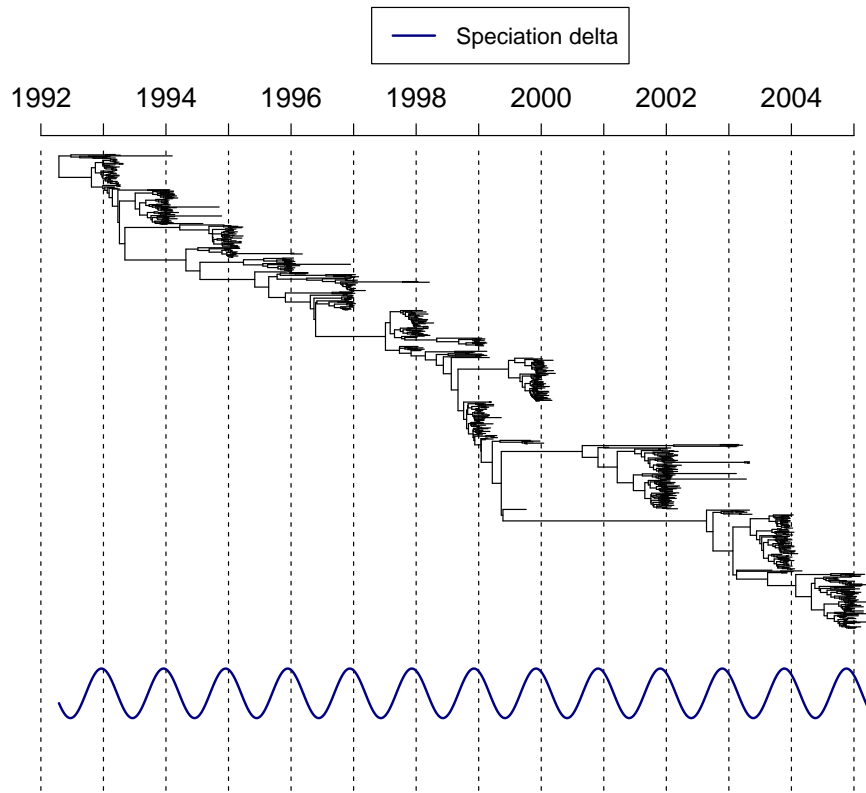
(a)

NA phylogeny



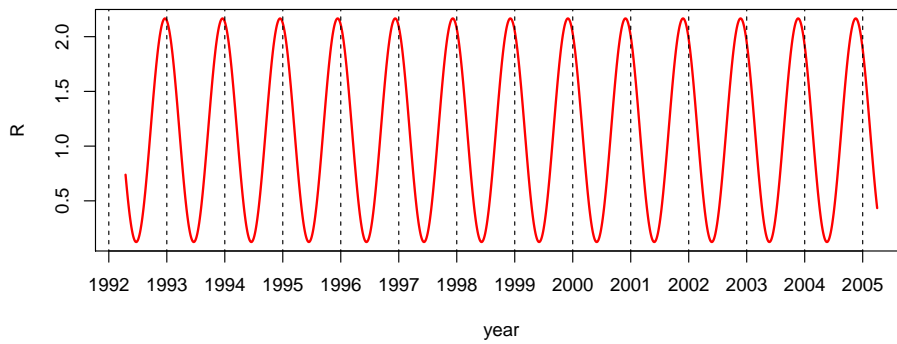
(b)

Figure B.1: The best fitted model for the NA phylogeny was the periodic speciation model. (a) The phylogeny and the best fitted periodic rate. (b) Estimated $R(t)$ from the best fitted model



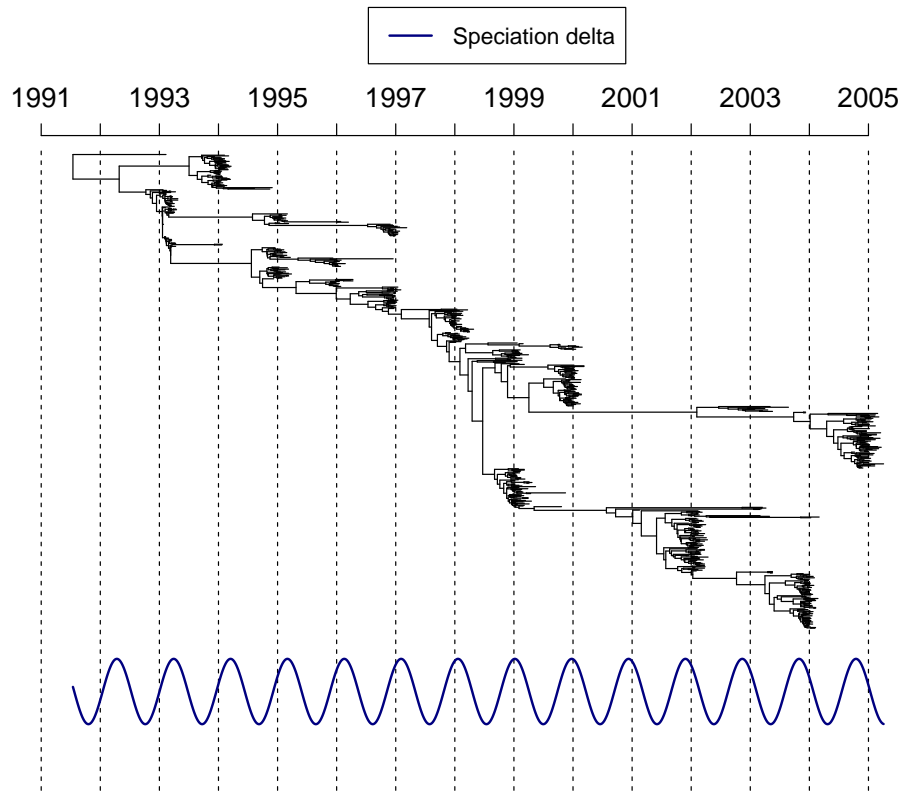
(a)

HA phylogeny



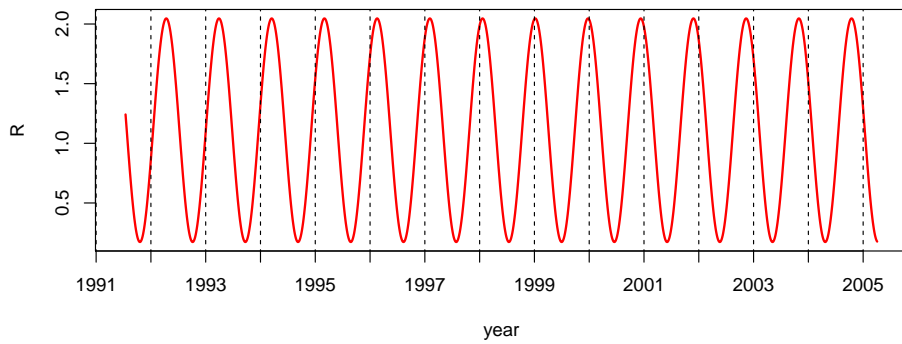
(b)

Figure B.2: The best fitted model for the HA phylogeny was the periodic speciation model with a shift. (a) The phylogeny and the best fitted periodic rate. (b) Estimated $R(t)$ from the best fitted model



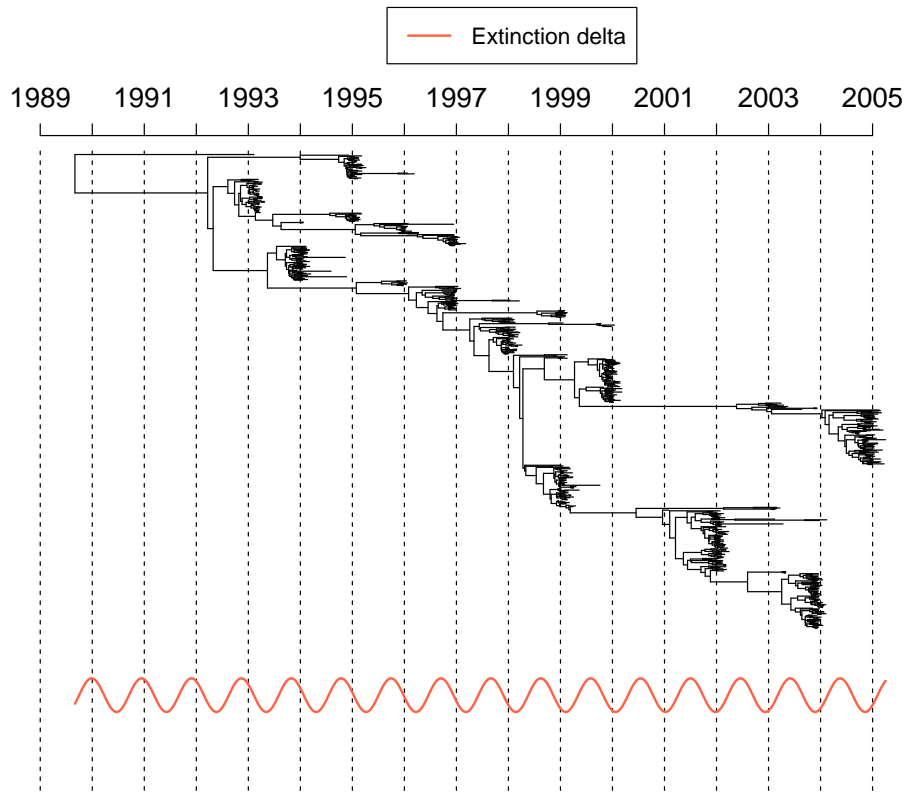
(a)

PA phylogeny



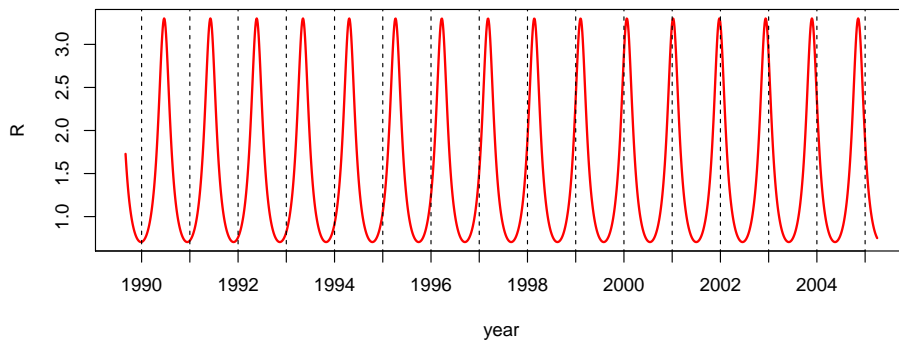
(b)

Figure B.3: The best fitted model for the PA phylogeny was the periodic speciation model with a shift. (a) The phylogeny and the best fitted periodic rate. (b) Estimated $R(t)$ from the best fitted model



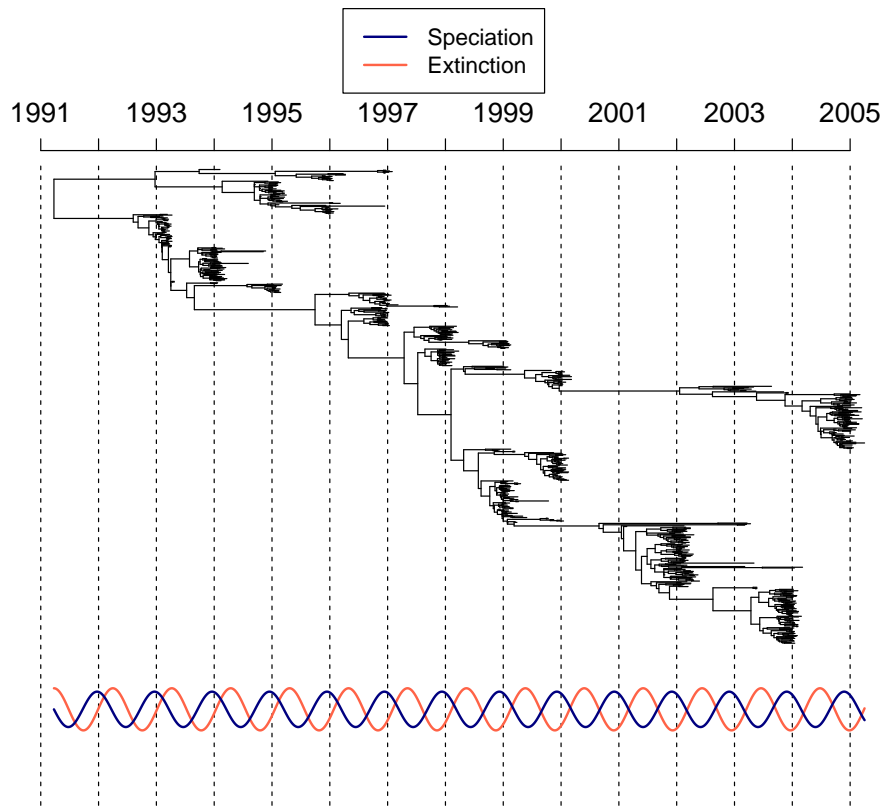
(a)

PB1 phylogeny



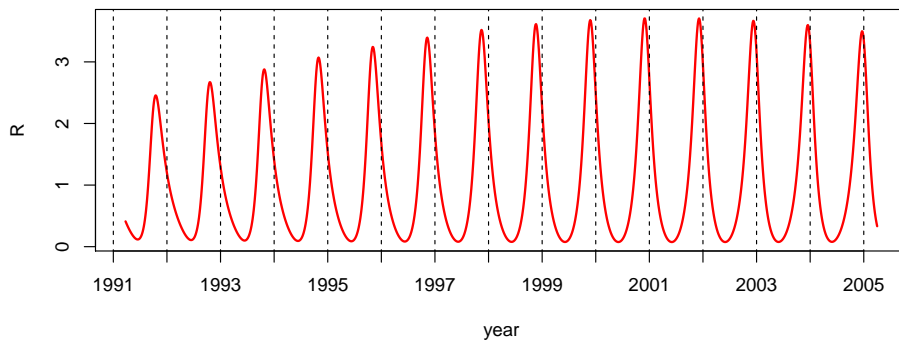
(b)

Figure B.4: The best fitted model for the PB1 phylogeny was the periodic speciation model with a shift. (a) The phylogeny and the best fitted periodic rate. (b) Estimated $R(t)$ from the best fitted model



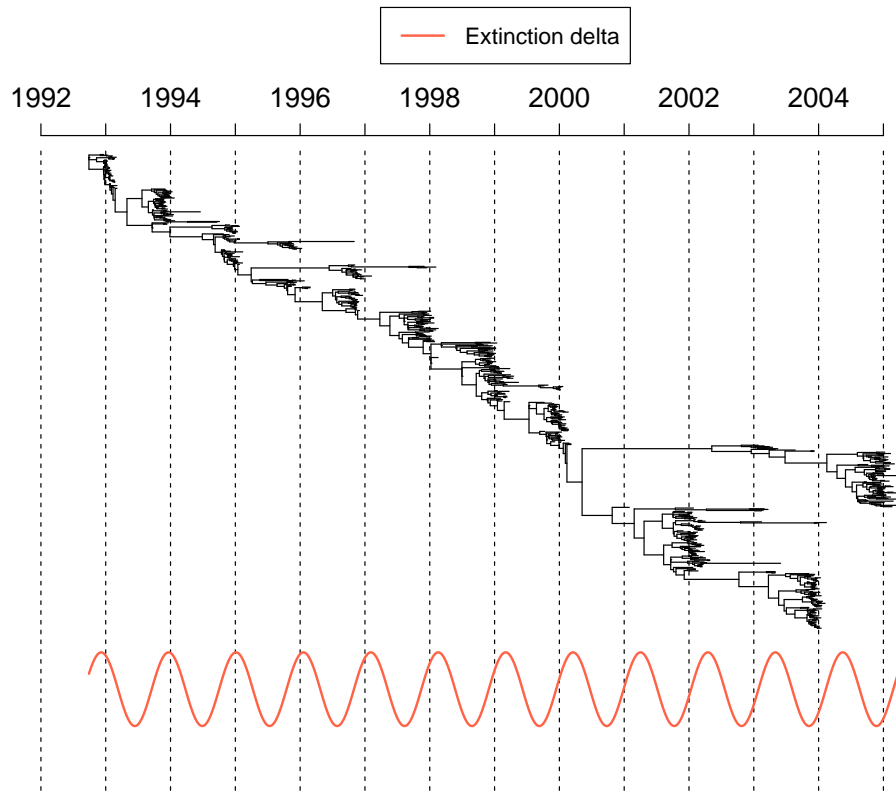
(a)

PB2 phylogeny



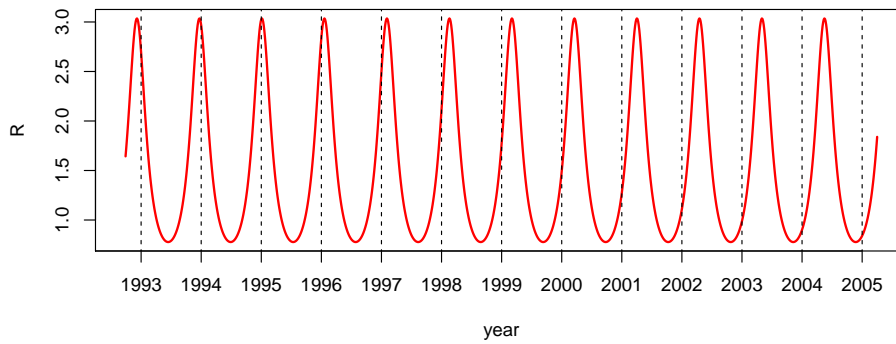
(b)

Figure B.5: The best fitted model for the PB2 phylogeny was the periodic speciation and extinction model. (a) The phylogeny and the best fitted periodic rate. (b) Estimated $R(t)$ from the best fitted model



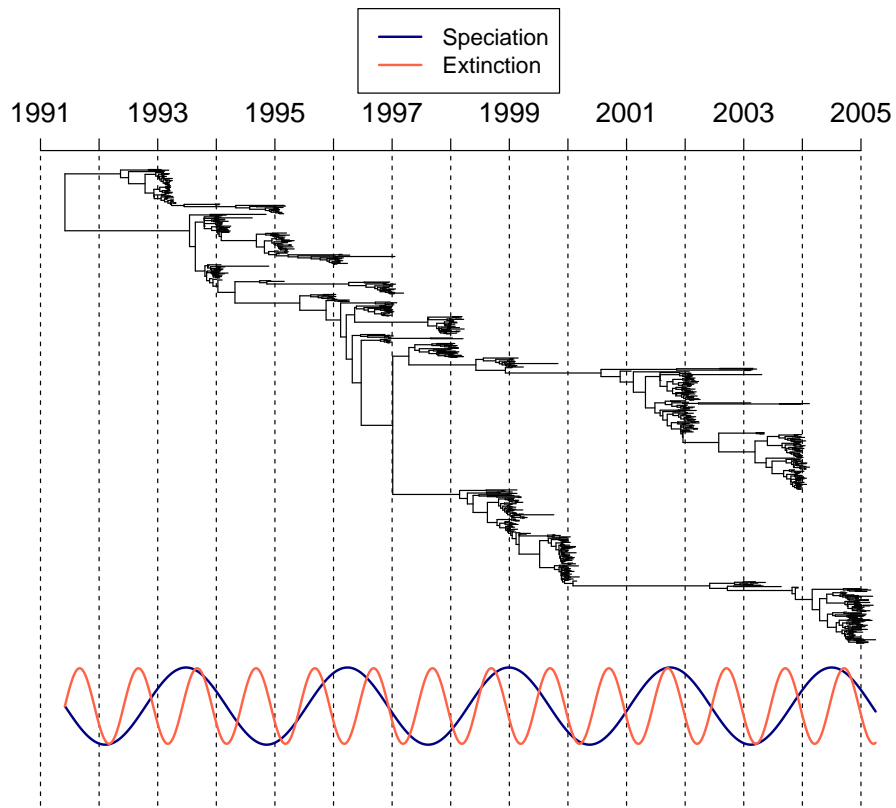
(a)

M12 phylogeny



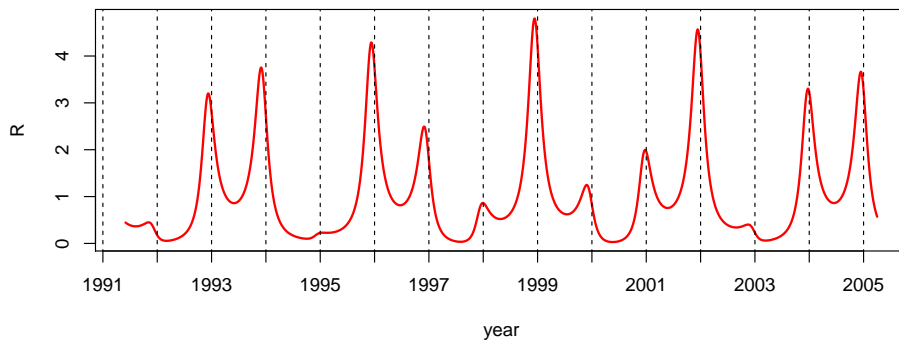
(b)

Figure B.6: The best fitted model for the M1/2 phylogeny was the periodic extinction model with a shift. (a) The phylogeny and the best fitted periodic rate. (b) Estimated $R(t)$ from the best fitted model



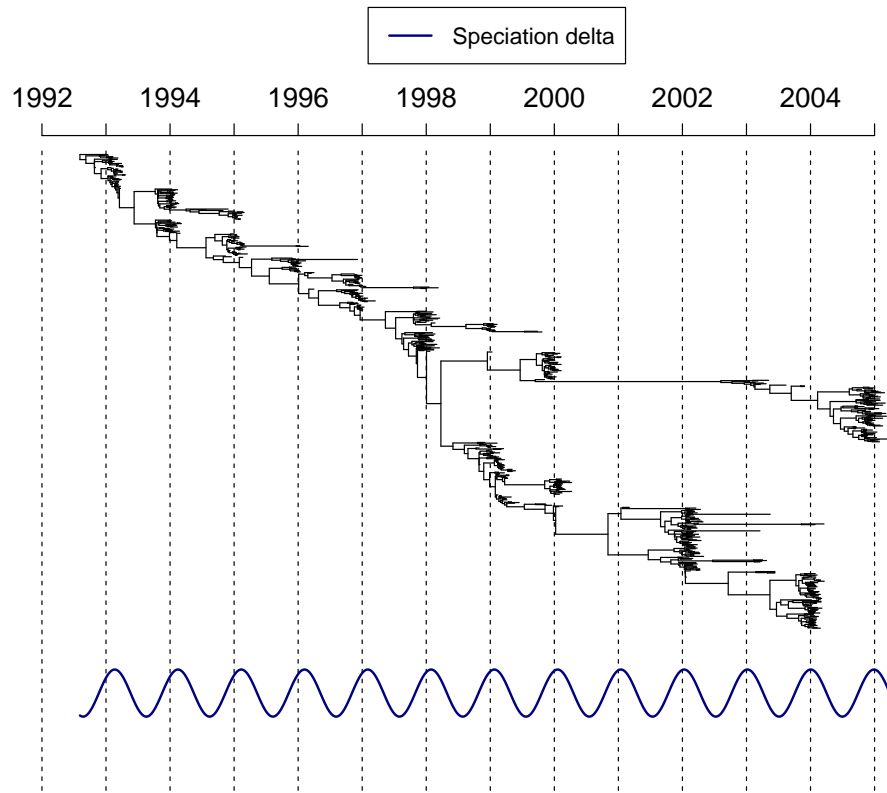
(a)

NP phylogeny



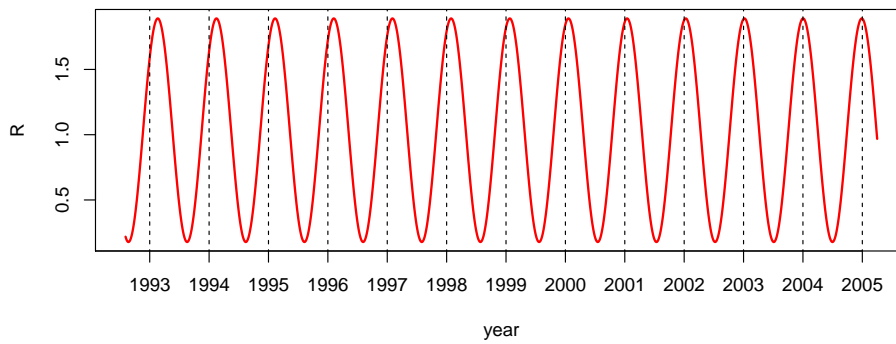
(b)

Figure B.7: The best fitted model for the NP phylogeny was the periodic speciation and extinction model. (a) The phylogeny and the best fitted periodic rate. (b) Estimated $R(t)$ from the best fitted model



(a)

Ns12 phylogeny



(b)

Figure B.8: The best fitted model for the Ns1/2 phylogeny was the periodic speciation model. (a) The phylogeny and the best fitted periodic rate. (b) Estimated $R(t)$ from the best fitted model

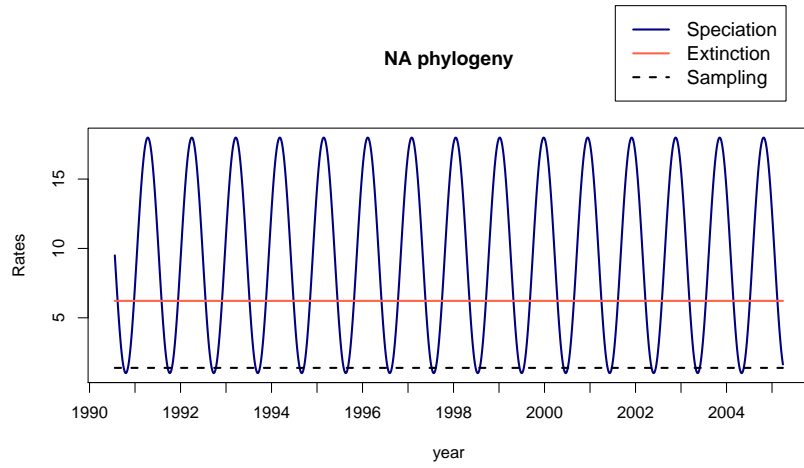


Figure B.9: NA phylogeny: Inferred rates for the best fitted model

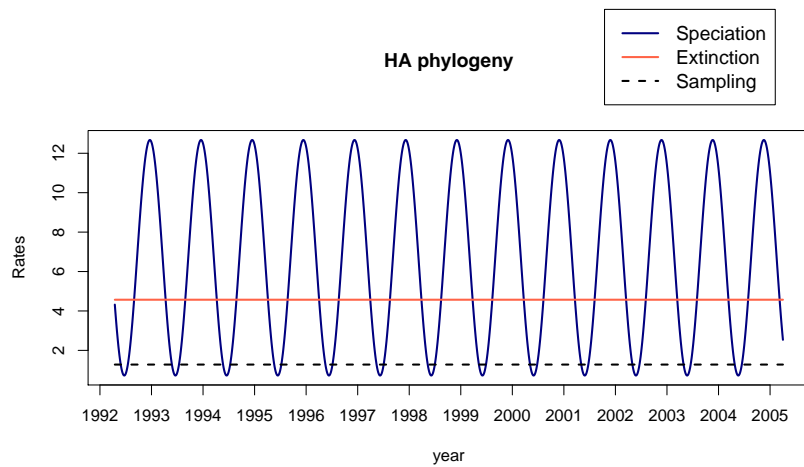


Figure B.10: HA phylogeny: Inferred rates for the best fitted model

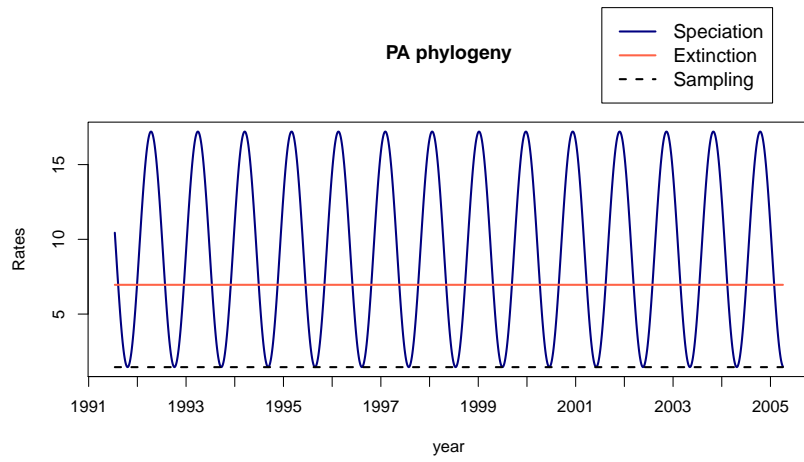


Figure B.11: PA phylogeny: Inferred rates for the best fitted model

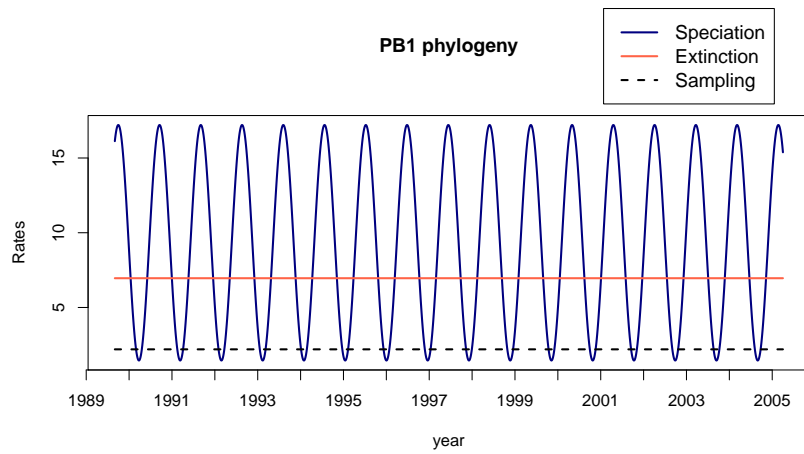


Figure B.12: PB1 phylogeny: Inferred rates for the best fitted model

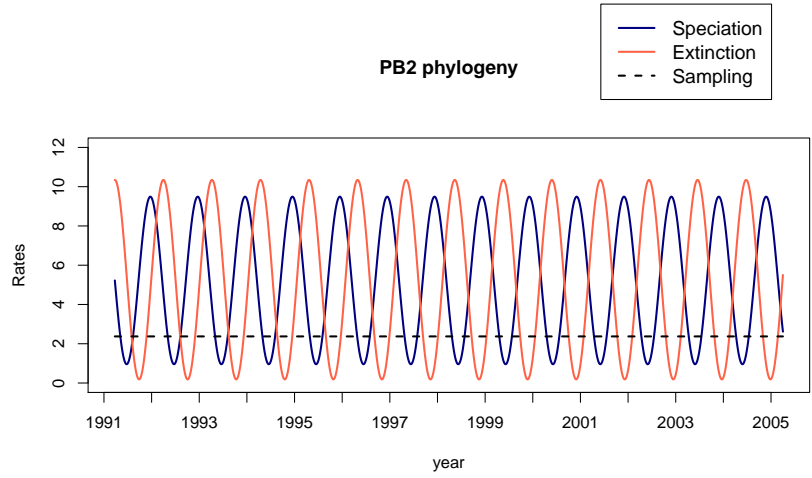


Figure B.13: PB2 phylogeny: Inferred rates for the best fitted model

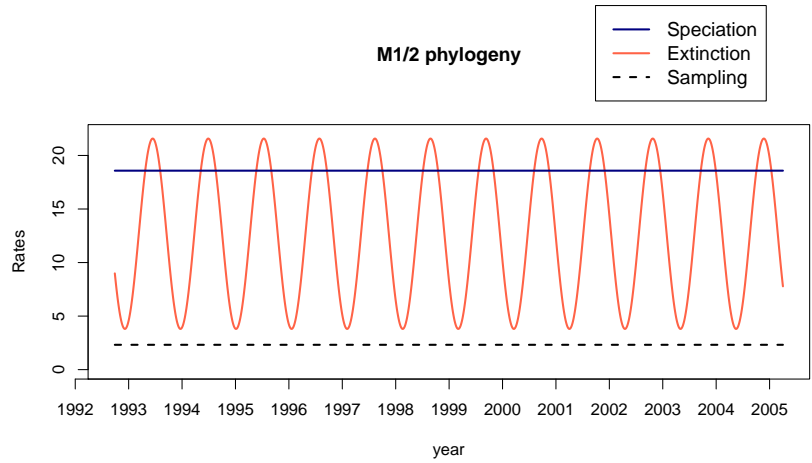


Figure B.14: M1/2 phylogeny: Inferred rates for the best fitted model

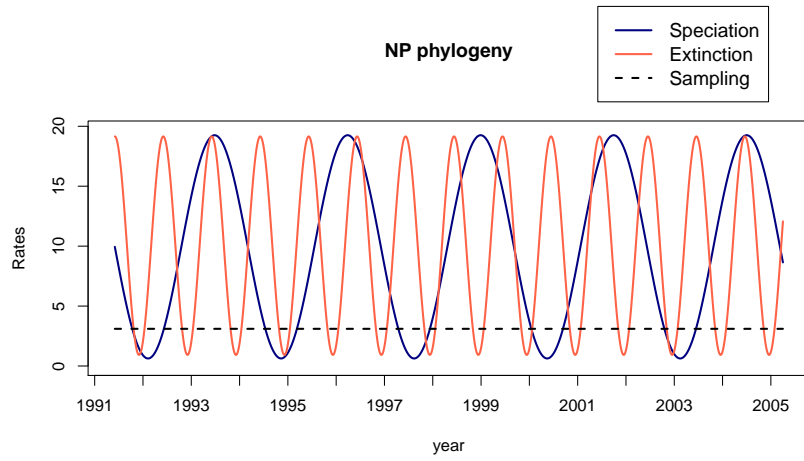


Figure B.15: NP phylogeny: Inferred rates for the best fitted model

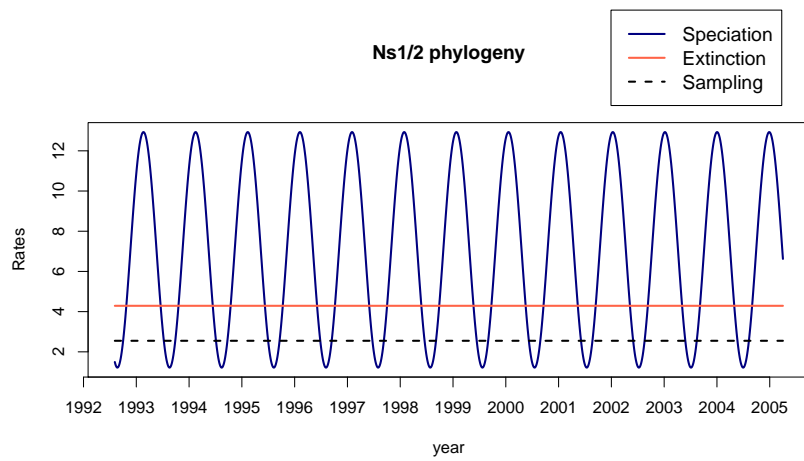


Figure B.16: Ns1/2 phylogeny: Inferred rates for the best fitted model

C Simulations

Table C.1: We simulated 500 trees with true values $\lambda = 3$, $\mu = 1.89$, $\psi = 0.81$ and $R_0 = 1.11$. On these we made inference assuming constant rates. In this table we show the result of the 100 first trees

Iter	$\hat{\lambda}$	$\hat{\mu}$	$\hat{\psi}$	\hat{R}_0	Iter	$\hat{\lambda}$	$\hat{\mu}$	$\hat{\psi}$	\hat{R}_0
1	4.4116	3.3667	0.6602	1.0955	51	5.1081	3.3667	0.4967	1.0946
2	2.62	1.2632	0.9491	1.1843	52	2.3716	1.2632	1.024	1.2419
3	2.3641	0.7178	1.1441	1.2697	53	3.5498	0.7178	0.7787	1.0769
4	3.2522	2.116	0.8042	1.1137	54	2.5553	2.116	1.0047	1.1229
5	1.9071	0.1371	1.4545	1.1983	55	3.7439	0.1371	0.7021	1.0656
6	1.783	0.1223	1.4317	1.1473	56	2.2324	0.1223	0.9635	1.2333
7	2.7359	1.41	0.9213	1.1735	57	2.2668	1.41	1.0504	1.0722
8	2.5499	1.1577	1.0984	1.1302	58	4.0213	1.1577	0.6579	1.0909
9	5.2775	4.3346	0.5669	1.0767	59	4.9289	4.3346	0.5389	1.0817
10	3.5598	2.4142	0.7509	1.1247	60	2.9556	2.4142	0.8763	1.0645
11	3.1906	2.1881	0.7967	1.069	61	3.2926	2.1881	0.7555	1.1888
12	3.3201	2.1301	0.826	1.1231	62	4.796	2.1301	0.5344	1.0843
13	1.8569	0.2028	1.4655	1.113	63	2.4591	0.2028	1.0358	1.1262
14	3.0484	1.8862	0.789	1.1395	64	2.5089	1.8862	1.0124	1.1382
15	5.1083	4.3019	0.5131	1.0609	65	3.5356	4.3019	0.7407	1.1198
16	2.0756	0.4637	1.2998	1.177	66	2.6947	0.4637	1.0363	1.1625
17	3.8189	2.8263	0.6388	1.1021	67	2.5671	2.8263	1.0821	1.1197
18	3.1833	2.1453	0.7524	1.0986	68	3.8158	2.1453	0.7288	1.0795
19	2.2375	0.8556	1.1304	1.1266	69	2.4229	0.8556	1.0246	1.234
20	3.4687	2.3977	0.7482	1.1026	70	1.9477	2.3977	1.3126	1.2486
21	2.7353	1.4356	1.0051	1.1207	71	3.334	1.4356	0.8273	1.2266
22	3.2988	2.0993	0.7896	1.1419	72	3.5498	2.0993	0.7058	1.1037
23	1.9784	0.2687	1.3907	1.1922	73	2.5771	0.2687	1.0597	1.1868
24	3.4438	2.2947	0.8213	1.1052	74	3.0584	2.2947	0.7884	1.0547
25	4.0752	3.1271	0.6703	1.0731	75	3.7839	3.1271	0.6585	1.061
26	2.4202	1.0667	1.0028	1.1694	76	2.0351	1.0667	1.4552	1.2636
27	3.5046	2.5518	0.7372	1.0656	77	2.4721	2.5518	1.0706	1.2176
28	2.9358	1.8247	0.8643	1.0918	78	4.4127	1.8247	0.5948	1.0991
29	3.0217	1.9916	0.9303	1.0342	79	4.5905	1.9916	0.5428	1.0861
30	3.4275	2.4223	0.7248	1.0891	80	2.5179	2.4223	1.1535	1.1134
31	5.3979	4.5821	0.5005	1.062	81	4.8489	4.5821	0.4816	1.1261
32	1.8272	0.1349	1.4377	1.1619	82	3.4408	0.1349	0.7487	1.0995
33	3.1804	1.9528	0.8708	1.1263	83	2.0872	1.9528	1.193	1.1619
34	3.6239	2.6169	0.7994	1.0608	84	3.0545	2.6169	1.0062	1.0943
35	1.7617	0.0139	1.4436	1.2087	85	2.387	0.0139	1.0098	1.1267
36	1.9585	0.422	1.3546	1.1024	86	3.3907	0.422	0.7738	1.1408
37	2.7936	1.372	0.919	1.2194	87	3.1647	1.372	0.8238	1.1433
38	3.9595	3.0081	0.7098	1.065	88	4.6003	3.0081	0.5339	1.0647
39	2.5435	1.149	1.1501	1.1063	89	2.24	1.149	1.0824	1.2339
40	2.4678	1.1398	1.0871	1.1082	90	2.6277	1.1398	1.027	1.0923
41	1.9628	0.3499	1.3458	1.1575	91	2.631	0.3499	0.936	1.2328
42	4.1376	3.2297	0.6096	1.0777	92	2.8678	3.2297	0.8732	1.0996
43	1.8896	0.1577	1.395	1.217	93	2.6463	0.1577	0.9028	1.1387
44	2.5736	1.1461	1.1684	1.1119	94	1.8751	1.1461	1.2572	1.3056
45	2.5894	1.1563	0.9512	1.2286	95	1.9278	1.1563	1.3732	1.2798
46	2.4033	0.9662	1.0302	1.2039	96	4.3623	0.9662	0.5998	1.0718
47	3.4386	2.2862	0.8878	1.0834	97	1.8693	2.2862	1.2031	1.2249
48	4.3241	3.3447	0.6135	1.0924	98	1.9013	3.3447	1.3229	1.1342
49	2.645	1.3031	0.9921	1.1524	99	4.8466	1.3031	0.5113	1.0622
50	1.8536	0.0393	1.4031	1.2851	100	2.4598	0.0393	0.9615	1.0966

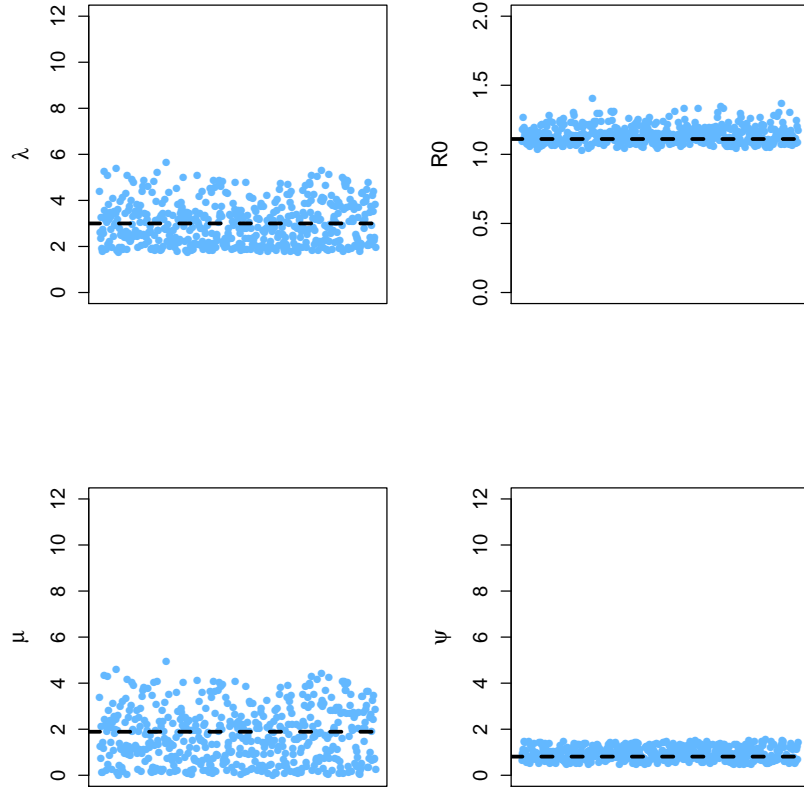


Figure C.1: Inferred constant rates for 500 trees simulated with $\lambda = 3$, $\mu = 1.89$, $\psi = 0.81$ and $R_0 = 1.111$.

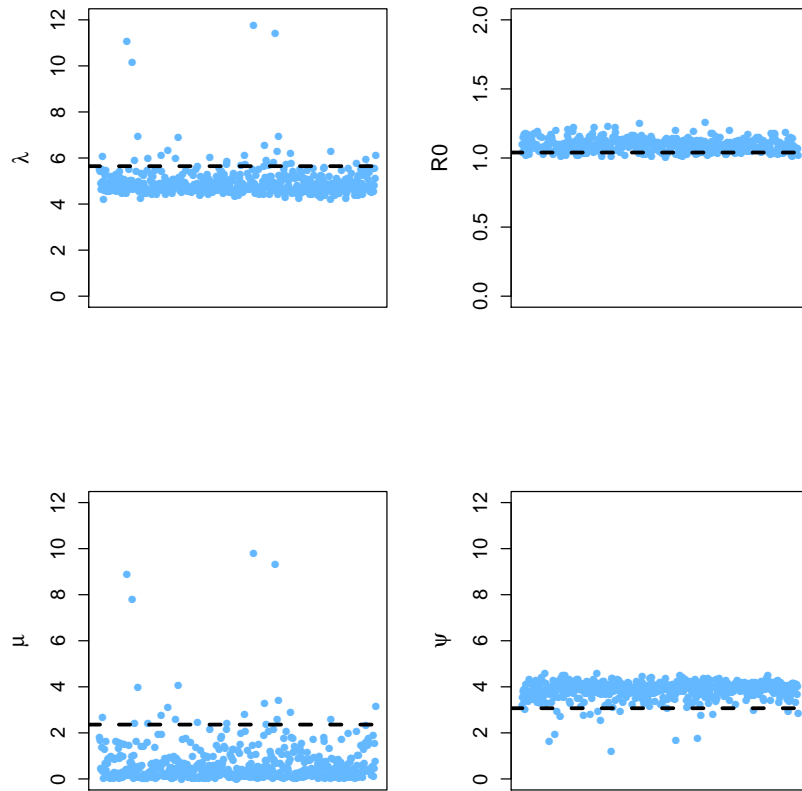


Figure C.2: Inferred constant rates for 500 trees simulated with $\lambda = 5.6451$, $\mu = 2.3590$, $\psi = 3.0703$ and $R_0 = 1.0397$. The death rate μ is underestimated but instead the sampling rate ψ is overestimated.

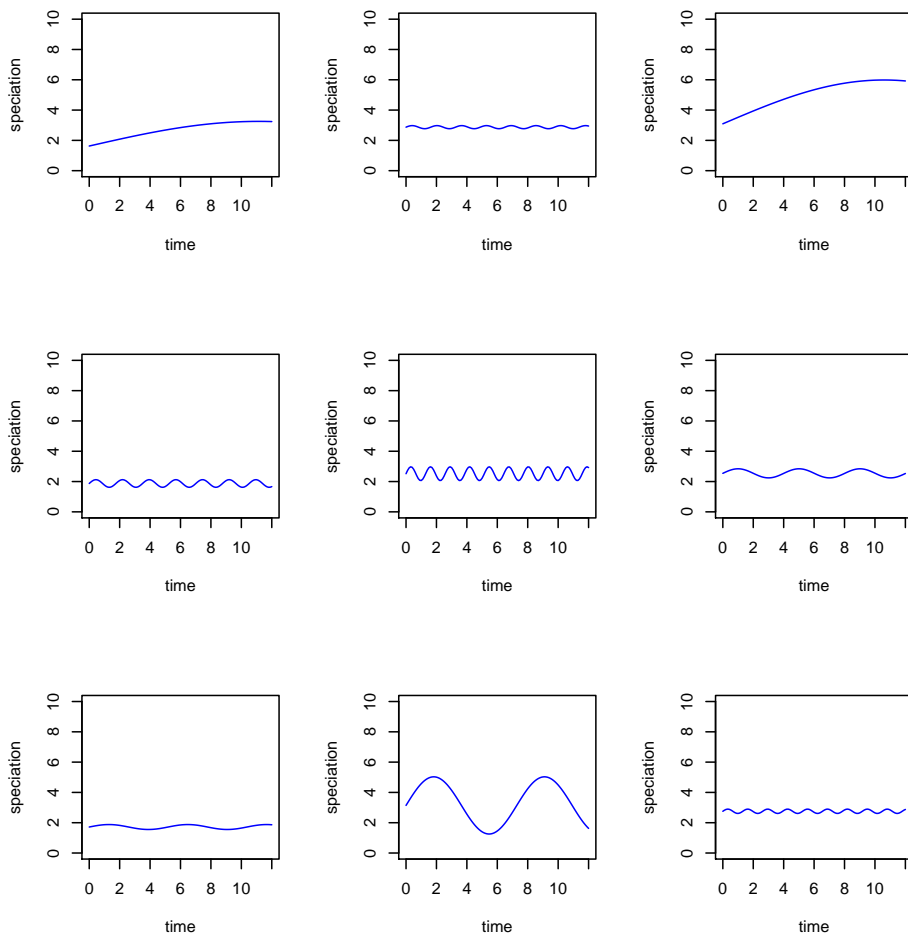


Figure C.3: Example of inferred periodic speciation rate when underlying tree is simulated under constant rates. The true speciation rate was set to $\lambda = 3$

Table C.2: Log-Likelihood for the 100 trees simulated with constant rates when inference was done under a constant rate model and periodic speciation rate model

iter	constant	periodic	iter	constant	periodic
1	-424.6458	-478.5449	51	-561.8902	-588.8464
2	-494.8615	-536.9854	52	-449.8783	-485.7118
3	-431.8928	-526.1300	53	-530.7630	-549.8356
4	-501.2279	-535.2022	54	-537.1609	-579.6032
5	-470.3611	-512.8584	55	-510.3708	-549.0917
6	-546.8867	-566.4281	56	-492.1084	-523.3704
7	-493.5461	-538.8127	57	-532.9484	-572.3753
8	-485.5151	-514.4469	58	-565.1363	-598.5613
9	-416.9010	-455.9329	59	-550.1233	-596.0942
10	-474.7734	-577.7230	60	-514.8717	-543.5579
11	-513.4897	-512.36	61	-498.4026	-498.4044
12	-515.2188	-1277.659	62	-515.3388	-515.3833
13	-468.3095	-662.7967	63	-484.0128	-483.8069
14	-520.5012	-519.3073	64	-486.9801	-485.0381
15	-470.834	-330.9536	65	-492.692	-492.2102
16	-499.1173	-499.0682	66	-557.7946	-556.7082
17	-553.2144	-553.2	67	-534.7668	-534.7309
18	-500.1261	-494.9736	68	-472.7075	-472.7512
19	-481.7441	-480.3516	69	-472.818	-472.4298
20	-504.0929	-503.7562	70	-430.1861	-429.1955
21	-532.5732	-532.5835	71	-473.5103	-473.4741
22	-544.9967	-544.5203	72	-479.127	-477.7797
23	-533.9102	-531.6386	73	-538.7356	-538.1232
24	-480.9457	-480.9406	74	-509.9671	-509.9452
25	-552.6558	-552.6393	75	-475.9463	-935.8914
26	-462.8317	-1174.369	76	-491.735	-621.0172
27	-477.6301	-474.8609	77	-485.7916	-1290.464
28	-519.1921	-518.2634	78	-491.6835	-489.8467
29	-466.3189	-465.7669	79	-480.4546	-480.4979
30	-556.9011	-556.4195	80	-512.2182	-511.1173
31	-534.8398	-534.8441	81	-475.137	-473.5378
32	-445.6848	-445.2923	82	-451.6511	-451.174
33	-475.4621	-475.4899	83	-476.4758	-475.0101
34	-505.5676	-504.8346	84	-454.3364	-443.5789
35	-450.4424	-450.5132	85	-481.5856	-480.2423
36	-524.9533	-524.9836	86	-568.5732	-568.6622
37	-515.3678	-514.231	87	-536.2316	-530.9844
38	-532.9167	-531.298	88	-541.8727	-541.9178
39	-582.5886	-582.4839	89	-524.2411	-1119.427
40	-519.4776	-514.4185	90	-577.61	-577.5678
41	-547.4655	-547.0376	91	-522.4404	-522.0645
42	-481.8287	-481.8398	92	-485.7611	-485.4511
43	-517.9969	-1208.227	93	-597.731	-543.6589
44	-563.2396	-562.9696	94	-510.262	-510.1174
45	-509.8942	-507.7519	95	-584.5472	-1204.523
46	-505.9405	-502.6053	96	-540.4654	-540.2616
47	-512.9387	-512.9528	97	-450.2732	-450.3064
48	-577.61	-577.5678	98	-514.7732	-514.4831
49	-563.7595	-558.3072	99	-542.9345	-542.7739
50	-472.9956	-472.8834	100	-505.2472	-226.313

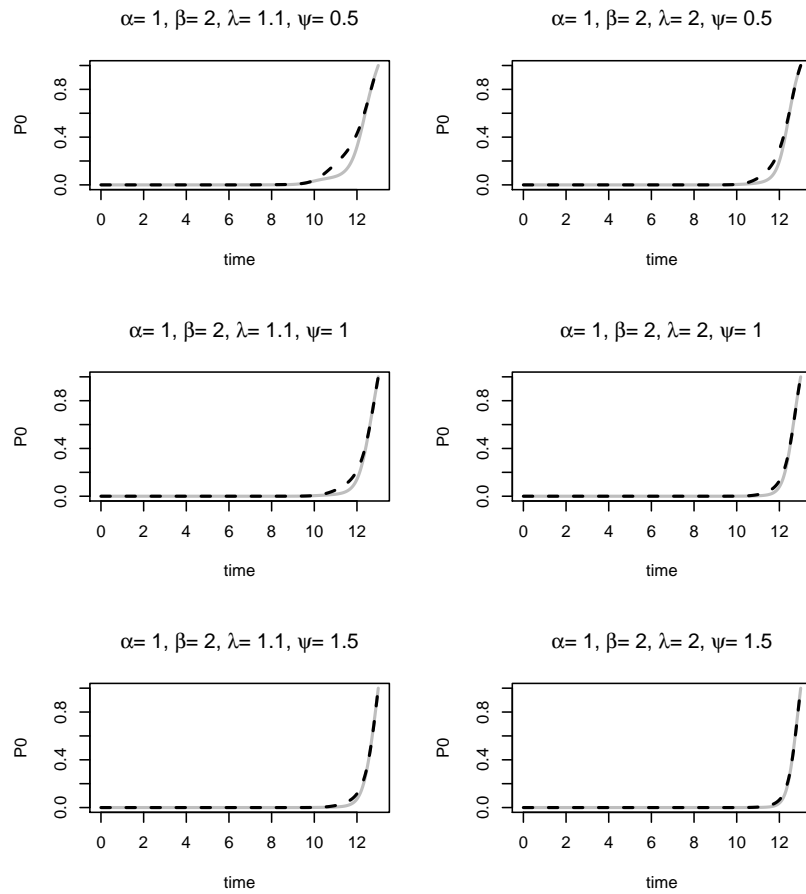


Figure C.4: Plot of numerical and analytical solution to P_0^S when $\mu(t) = 0$. The grey line is the numerical solution and the black line are the analytical solution

D R code

```
library(deSolve)

# Equations for calculating the likelihood.
# Assuming we have a vector with all times of speciation events: nodes,
# and a vector with all sampling times: tip.

ZERO_CUTOFF <- 1E-07

##### Constant speciation and constant extinction #####
#           lambda(t) = lambda           #
#           mu(t)    = mu               #
#####

#derivative of P0, to be thrown in in ode
dP0<- function(t, y, parms){
  lambda <- parms[1]
  mu     <- parms[2]
  psi    <- parms[3]
  list(mu-(lambda+mu+psi)*y+lambda*(y^2))
}

#solve the differential equation
diff2<-function(lambda,mu,psi,time){
  yini <- c(y=1)
  P0d <- ode(y=yini,func=dP0,times=time,parms=c(lambda,mu,psi))
  return(P0d)
}

# create values to be used to create a function for P0,
#the amount of values decides how accurate the function becomes

P0_allf<-function(lambda,mu,psi,T){
  times <- seq(0,T,by=0.0005)
  res   <- diff2(lambda,mu,psi,times,T)
  dtimes<- times
  dres  <- rev(res[,2])
  return(list(dtimes=dtimes,dres=dres))
}
```

```

# calculating P1 with integrate
P1_all<-function(lambda,mu,psi,times,res,t1,T){
  s      <- length(t1)
  f.P0   <- approxfun(times,res)
  integrand <- function(x){
    area   <- lambda + mu + psi - 2*lambda*f.P0(x)
    return(area)
  }
  integrandv<- Vectorize(integrand,"x")
  res <- 1
  if(s==1){
    if ( (T-t1) > ZERO_CUTOFF ) {
      res <-exp(-integrate(integrandv,lower=t1,upper=T)$value)
    }
  }
  if(s>1){
    res  <- c(1:s)
    dummy <- mapply(integrate, lower=t1,upper=T, MoreArgs=list(f=integrandv))
    for(l in 1:s) { res[l]<-exp(-dummy[l,][[1]])}
  }
  return(res)
}

# calculating P1 with differential equations instead of integrate
diffP1<-function(lambda,mu,psi,times,res,time,T){
  f.P0   <-approxfun(times,res)
  integrand <-function(x){
    area <- lambda + mu + psi - 2*lambda*f.P0(x)
    return(area)
  }
  integrandv <-Vectorize(integrand,"x")
  dP1 <- function(x, y, parms){
    lambda <- parms[1]
    mu     <- parms[2]
    psi    <- parms[3]
    list(lambda + mu + psi - 2*lambda*f.P0(x))
  }

  yin  <- integrate(integrandv,lower=0,upper=T)$value
  yini <- c(y=yin)
  P1d  <- ode(y=yini,func=dP1,times=time,parms=c(lambda,mu,psi))
  res  <- P1d[,2]
  times <- P1d[,1]
  pos  <- which(is.na(res))
  if((length(pos)) > 0){
    for(i in 1:(length(pos))){
      yt <-exp(-integrate(integrandv,lower=(times[(pos[i]-1)]),upper=times[pos[i]])$value)
      dum <- res[(pos[i]-1)]-log(yt)
      res[pos[i]] <- dum
    }
  }
  res <- res[length(res)]-res
  return(exp(-res))
}

```



```

}

# density of constant birth and death model with time varying rates
# and sampling through time

dBirthDeathdiff<- function(node,tip,birth,death,psi){
  node <- sort(node)
  tip <- sort(tip)
  nodeAndTip <- sort(c(node,tip))
  N <- length(tip)
  T <- tip[N]
  tmp <- PO_allf(birth,death,psi,T)

  all_times <- tmp$dtimes
  PO_res <- tmp$dres
  P1 <- diffP1(birth,death,psi,all_times,PO_res,nodeAndTip,T)
  fP1 <- approxfun(nodeAndTip,P1)
  pr <- (N-1) * log(birth) + N * log(psi)
  pr <- pr + sum(log(fP1(node)))-sum(log(fP1(tip)))

  #we condition on at least on sampled species
  pr <- pr - log(1-all_res[1])
  return(pr)
}

dBirthDeath <- function(node,tip,birth,death,psi){
  node <- sort(node)
  tip <- sort(tip)
  N <- length(tip)
  T <- tip[N]
  tmp <- PO_allf(birth,death,psi,T)

  all_times <- tmp$dtimes
  PO_res <- tmp$dres
  pr <- (N-1) * log(birth) + N * log(psi)
  pr <- pr + sum(log(P1_all(birth,death,psi,all_times,PO_res,node,T)))
  pr <- pr - sum(log(P1_all(birth,death,psi,all_times,PO_res,tip,T)))

  #we condition on at least on sampled species
  pr <- pr - log(1-all_res[1])
  return(pr)
}

```

```

##### Periodic speciation and periodic extinction #####
#           lambda(t) = a*sin(b*t)+c           #
#           mu(t)    = a2*cos(b2*t)+c2        #
#####

#derivative of P0, to be used in ode
dP0 <- function(t, y, parms){
  a  <- parms[1]
  b  <- parms[2]
  c  <- parms[3]
  a2 <- parms[4]
  b2 <- parms[5]
  c2 <- parms[6]
  psi <- parms[7]
  list(a2*cos(b2*t)+c2-(a*sin(b*t)+c+a2*cos(b2*t)+c2+psi)*y+(a*sin(b*t)+c)*(y^2))
}

#solve the differential equation
diff2 <- function(a,b,c,a2,b2,c2,psi,time){
  yini <- c(y=1)
  P0d <- ode(y=yini,func=dP0,times=time,parms=c(a,b,c,a2,b2,c2,psi))
  return(P0d)
}

# create values to be used to create a function for P0,
# the amount of values decides how accurate the function becomes

P0_allf <- function(a,b,c,a2,b2,c2,psi,T){
  times <- seq(0,T,by=0.0005)
  res <- diff2(a,b,c,a2,b2,c2,psi,times,T)
  dtimes <- times
  dres <- rev(res[,2])
  return(list(dtimes=dtimes,dres=dres))
}

```

```

#function P1 using integrate.
#Slower than diffP1 which integrates using differential equations
P1_all <- function(a,b,c,a2,b2,c2,psi,times,res,t1,T){
  s      <- length(t1)
  f.P0   <- approxfun(times,res)
  integrand <- function(x){
    area <- (a*sin(b*x)+c) + a2*cos(b2*x)+c2 + psi - 2*(a*sin(b*x)+c)*f.P0(x)
    return(area)
  }
  integrandv <- Vectorize(integrand,"x")
  res      <- 1
  if(s==1){
    if ( (T-t1) > ZERO_CUTOFF ) {
      res <- exp(-integrate(integrandv,lower=t1,upper=T)$value)
    }
  }
  if(s>1){
    res <- c(1:s)
    dummy <- mapply(integrate, lower=t1,upper=T, MoreArgs=list(f=integrandv))
    for(l in 1:s){ res[l]<-exp(-dummy[l,][[1]])}
  }
  return(res)
}

#diffP1 give same result as P1_all but is faster.
#Sometimes some values is not calculated and
#we use the same approach as in P1_all and use the function
#integrate to get these.

diffP1<-function(a,b,c,a2,b2,c2,psi,times,res,time,T){
  f.P0   <- approxfun(times,res)
  integrand <- function(x){
    area <- (a*sin(b*x)+c) + a2*cos(b2*x)+c2 + psi - 2*(a*sin(b*x)+c)*f.P0(x)
    return(area)
  }
  integrandv <- Vectorize(integrand,"x")
  dP1<- function(x, y, parms){
    a <- parms[1]
    b <- parms[2]
    c <- parms[3]
    a2 <- parms[4]
    b2 <- parms[5]
    c2 <- parms[6]
    psi <- parms[7]
    list((a*sin(b*x)+c) + a2*cos(b2*x)+c2 + psi - 2*(a*sin(b*x)+c)*f.P0(x))
  }

  yin <- integrate(integrandv,lower=0,upper=T)$value
  yini <- c(y=yin)
  P1d <- ode(y=yini,func=dP1,times=time,parms=c(a,b,c,a2,b2,c2,psi))
  res <- P1d[,2]
  times <- P1d[,1]

```

```

pos  <- which(is.na(res))
if((length(pos)) > 0){
  for(i in 1:(length(pos))){
    yt <-exp(-integrate(integrandv,lower=(times[(pos[i]-1)]),upper=times[pos[i]])$value)
    dum <-res[(pos[i]-1)]-log(yt)
    res[pos[i]]<-dum
  }
}
res <- res[length(res)]-res
return(exp(-res))
}

#calculated the likelihood
dBirthDeathdiff <- function(node,tip,a,b,c,a2,b2,c2,psi){
  node <- sort(node)
  tip <- sort(tip)
  nodeAndTip <- sort(c(node,tip))
  #get the number of species in the phylogeny
  N <- length(tip)
  #get the age of the phylogeny
  T <- tip[N]
  tmp <- P0_allf(a,b,c,a2,b2,c2,psi,T)

  all_times <- tmp$dtimes
  P0_res <- tmp$dres
  P1 <- diffP1(a,b,c,a2,b2,c2,psi,all_times,P0_res,nodeAndTip,T)
  fP1 <- approxfun(nodeAndTip,P1)

  pr <- N * log(psi)
  pr <- pr + sum(log(fP1(node)))+sum(log((a*sin(b*node)+c)))
  pr <- pr - sum(log(fP1(tip)))
  #we condition on at least one sampled lineage
  pr <- pr - log(1-all_res[1])
  return(pr)
}

```