APPENDIX

₆₄₀

₁ This online Appendix presents the detailed derivation of the model used by

₂ Andréoletti, Zwaans et al, as well as supplementary results and figures. We extend results

₃ of Gupta et al. (2020) and Manceau et al. (2021) to piecewise-constant parameters,

₄ describe our implementation in the RevBayes software, and give detailed information on

₅ all priors used for simulation or inference in our analyses.

A – Method extension to piecewise-constant parameters

Notation and outline of the general strategy

₈ We first recall in Figure S1 the notation that we introduced in the main text with

₉ the three different sampling (ψ-sampling for sampling of fossils with inclusion in the tree,

₁₀ ω-sampling for occurrences and ρ-sampling at present).

₁₁ To compute the likelihood of $(\mathcal{T}, \mathcal{O})$ under this process, we slice horizontally our

₁₂ observations and perform a breadth-first traversal of these. We thus introduce now,

$$\mathcal{T}_t^\uparrow := \text{the tree } \mathcal{T} \text{ cut at time } t$$

$$\mathcal{T}_t^\downarrow := \text{the collection of trees (or forest) obtained by cutting } \mathcal{T}$$

$$\text{at time } t, \text{ and considering all subtrees descending from cut lineages}$$

$$k_t := \text{number of sampled lineages in } \mathcal{T} \text{ at time } t$$

$$\mathcal{O}_t^\uparrow := \mathcal{O}_{|(t,+\infty)}$$

$$\mathcal{O}_t^\downarrow := \mathcal{O}_{|(0,t)}$$

₁₃ We can now recall the definition of our two key probability densities,

$$\forall i \in \mathbb{N}, \quad L_t^{(i)} := \mathbb{P}(T_t^\downarrow, \mathcal{O}_t^\downarrow \mid I_t = k_t + i) \tag{S1}$$

$$\forall i \in \mathbb{N}, \quad M_t^{(i)} := \mathbb{P}(T_t^\uparrow, \mathcal{O}_t^\uparrow, I_t = k_t + i) \tag{S2}$$

₁₄ These probability densities have been introduced in Manceau et al. (2021) as a way

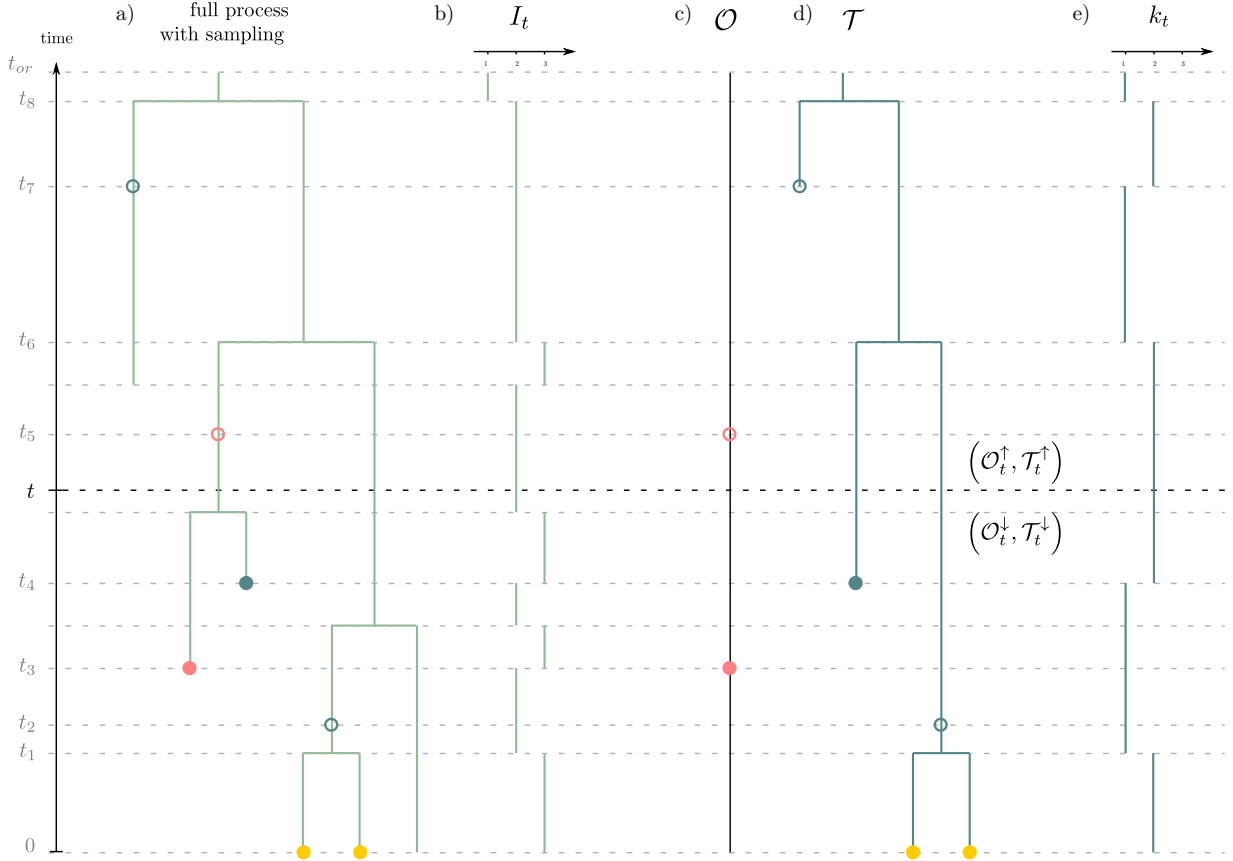₁₅ to target the probability distribution K_t of the total number of lineages given the data.

Figure S1. General setting of the method. a) the full process with sampling. Pink dots correspond to ω-sampling (sampling through time without sequencing), blue dots correspond to ψ-sampling (sampling through time with sequencing) and yellow dots correspond to ρ-sampling at present. Filled or unfilled dots correspond respectively to sampling with or without removal. b) Total number of lineages through time. c) Record of occurrences. d) Reconstructed tree spanning ψ- and ρ-samples. e) Number of lineages through time in the reconstructed tree (i.e. LTT plot).

Indeed,

$$
\begin{aligned}
K_t^{(i)} &:= \mathbb{P}(I_t = k_t + i \mid \mathcal{T}, \mathcal{O}) \\
&\propto \mathbb{P}(I_t = k_t + i, T_t^{\uparrow}, \mathcal{O}_t^{\uparrow}, T_t^{\downarrow}, \mathcal{O}_t^{\downarrow}) \\
&\propto \mathbb{P}(T_t^{\downarrow}, \mathcal{O}_t^{\downarrow} \mid I_t = k_t + i, T_t^{\uparrow}, \mathcal{O}_t^{\uparrow}) \mathbb{P}(I_t = k_t + i, T_t^{\uparrow}, \mathcal{O}_t^{\uparrow}) \\
&\propto L_t^{(i)} M_t^{(i)}
\end{aligned}
\tag{S3}
$$

The general strategy of the methods consists of (i) traversing the data backward in time to compute L_t; (ii) traversing the data forward in time to compute M_t; (iii) using the results to compute K_t. This scheme is illustrated in Figure S2.
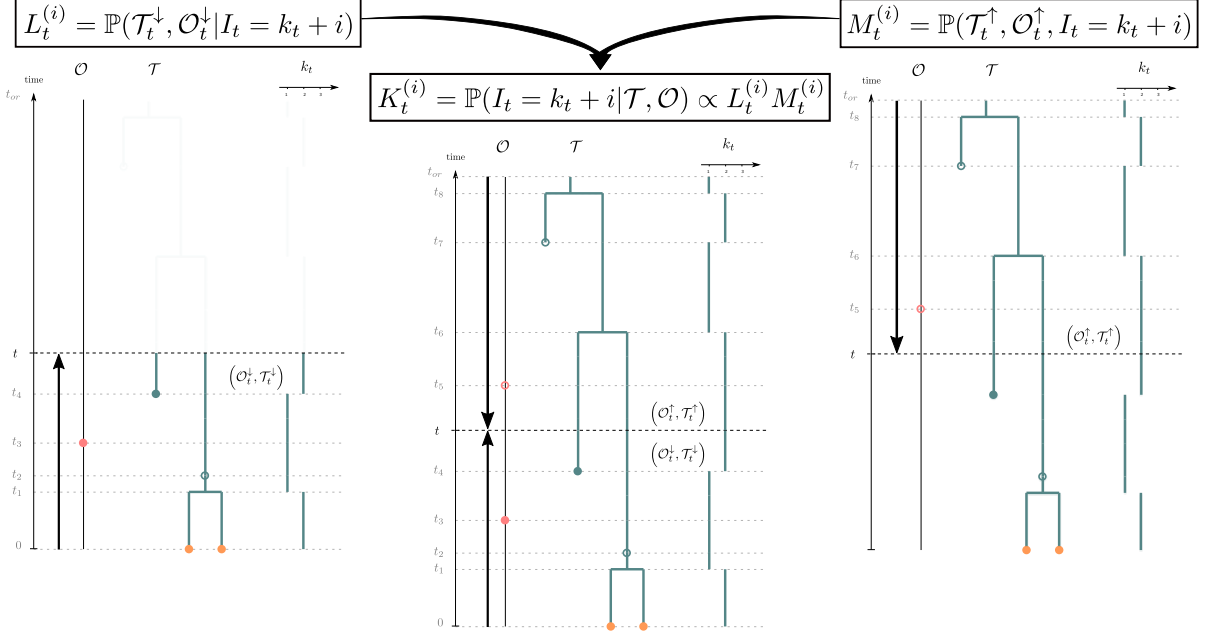
Figure S2. Inferring the posterior distribution of the number of lineages (K_t) in the OBDP. The probability distribution of the past number of lineages K_t is obtained at each time t by combining the quantity L_t obtained from the backward traversal algorithm (left) and the quantity M_t obtained from the forward traversal algorithm (right). See Table 1 for notations.

In the rest of this online Appendix section, we present the Master equations governing the evolution of these densities through time in a setup with piecewise-constant parameters.

Temporal setup for piecewise constant parameters

We partition time into two distinct units.

First, we define periods of time with no observations or sampling events, coined *epochs*, which allow for the basic derivation of Master equations of L_t and M_t. Epochs are delimited by all n *punctual events* times (i.e. branching and sampling events) in $\mathcal{O}$ and $\mathcal{T}$ pooled in an ordered list $(t_h)_{h=1}^n$. Epoch h is thus defined as the time interval (t_h, t_{h+1}).

Second, we account for all rate shift events, which define *constant rate time intervals*. If we have m such intervals, we pool all $m+1$ rate shift events in an ordered list $(\tau_l)_{l=0}^{m+1}$, where by convention we consider that $\tau_0 = 0$ and $\tau_{m+1} = t_{or}$. Rate time interval l

32   is defined as (τ_l, τ_{l+1}), with parameter set $(\lambda_l, \mu_l, \psi_l, \omega_l, r_l)$. We illustrate this setup in
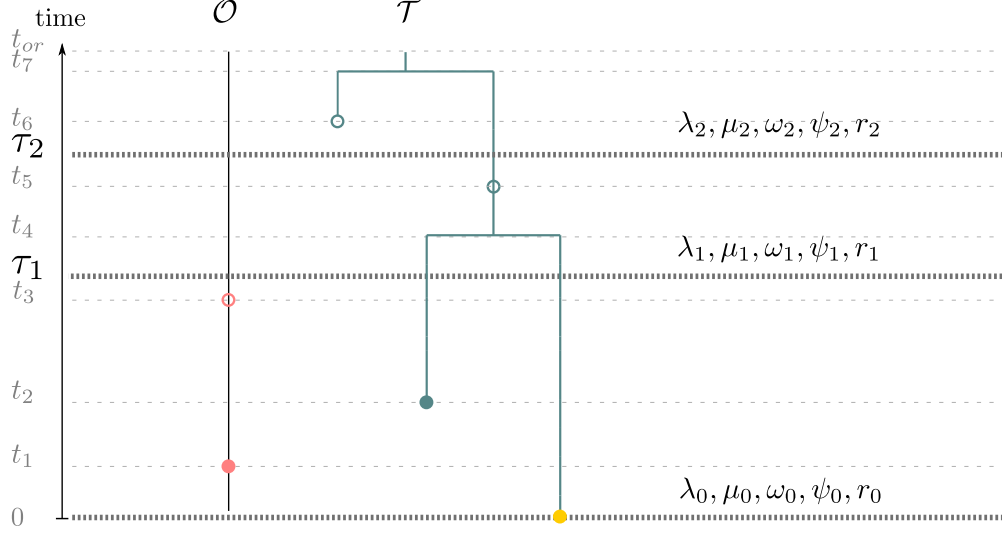
33   Figure S3 below.

Figure S3. Temporal setup of the method.

34   *Master equations governing Lt and Mt*

35   Probability densities L_t and M_t satisfy different Master equations obtained by

36   studying their evolution through time along any given epoch. These are ordinary

37   differential equations (ODE) that can be approximated numerically. Here, we assume

38    $\tau_l \leqslant t < \tau_{l+1}$ meaning that parameters have values $(\lambda_l, \mu_l, \psi_l, \omega_l, r_l)$.

39   First, we can initialize L_t and M_t respectively at present time 0 and at the time of

40   origin t_{or}. At present, ρ sampling of extant tips yields,

$$\forall i \in \mathbb{N}, \quad L_0^{(i)} = \rho^{k_0}(1-\rho)^i \tag{S4}$$

41   while at the time of origin, the process starts with only one lineage $k_{t_{or}} = 1$, which yields,

$$\forall i \in \mathbb{N}, \quad M_{t_{or}}^{(i)} = \mathbb{P}(I_{t_{or}} = 1 + i) = \mathbb{1}_{i=0} \tag{S5}$$

42   We now consider all events happening in an infinitesimal time step δt in the full

underlying process which do not result in observations or samplings. Three scenarios correspond to this case:

1. nothing happened with probability $(1 - \gamma_l(k+i)\delta t)$, where $\gamma_l = \lambda_l + \mu_l + \psi_l + \omega_l$

2. a birth event happened :

   (a) among the k sampled lineages in $T_t^{\downarrow}$, and it leads to an extinct or unsampled subtree to the left or to the right with probability $2\lambda_l k \delta t$

   (b) among the i other lineages with probability $\lambda_l i \delta t$.

3. a death event happened among the i particles, with probability $\mu_l i \delta t$

   We combine these to write, $\forall i \in \mathbb{N}$,

$$L_{t+\delta_t}^{(i)} = (1 - \gamma_l(k+i)\delta t)L_t^{(i)} + \lambda_l(2k+i)\delta t)L_t^{(i+1)} + \mu_l i \delta t)L_t^{(i-1)} \tag{S6}$$

   Letting $\delta t \to 0$ yields the following differential equation for L_t,

$$\forall i \in \mathbb{N}, \ \ L_0^{(i)} = \rho^{k_0}(1-\rho)^i \tag{S7}$$

$$\dot{L}_t^{(i)} = -\gamma_l(k+i)L_t^{(i)} + \lambda_l(2k+i)L_t^{(i+1)} + \mu_l i L_t^{(i-1)} \tag{S8}$$

   Similarly, M_t is the solution of the following ODE,

$$\forall i \in \mathbb{N}, \ \ M_{t_{or}}^{(i)} = \mathbb{P}(I_{t_{or}} = 1 + i) = \mathbb{1}_{i=0} \tag{S9}$$

$$\dot{M}_t^{(i)} = -\gamma_l(k+i)M_t^{(i)} + \lambda_l(2k+i-1)M_t^{(i-1)} + \mu_l(i+1)M_t^{(i+1)} \tag{S10}$$

Updates at punctual events

There are 6 types of punctual events in $\mathcal{T}$ and $\mathcal{O}$ that affect the probability densities M_t and L_t. These correspond to all different sampling options along $\mathcal{T}$ and $\mathcal{O}$ as illustrated in Figure S4. We denote as M_{t-} and L_{t-} the probability densities immediately prior to the event and M_{t+} and L_{t+} immediately after each event. We emphasise that the expressions differ when considering the process forward in time for M_t or backward in time, for L_t. These cases are the following :
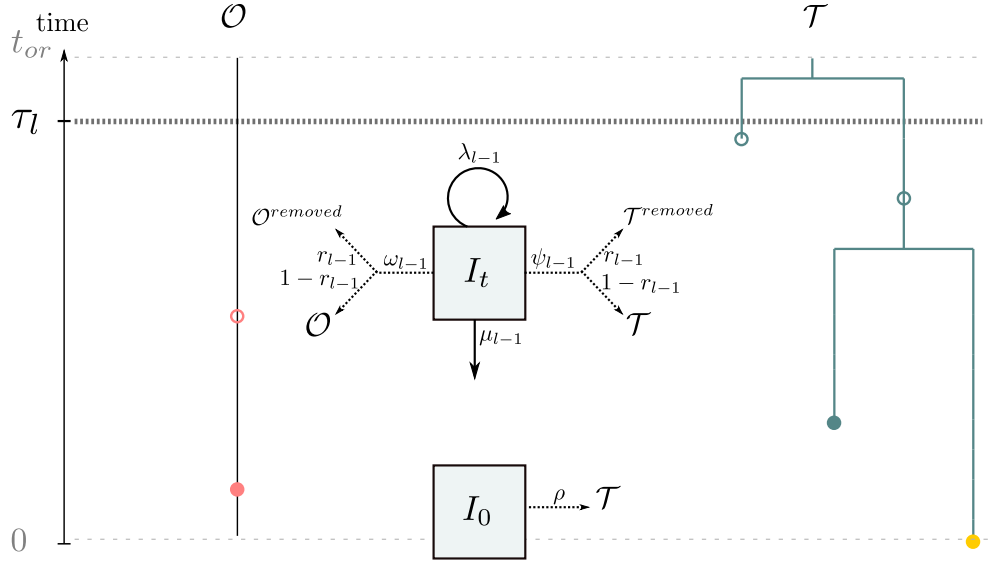
Figure S4. Updated sampling scheme of the method.

1. sampling of a leaf:

    (a) in $\mathcal{T}_t^{\downarrow}$, $L_{t^+}^{(i)} = \psi_l(1 - r_l)L_{t^-}^{(i+1)}$

    (b) in $\mathcal{T}_t^{\uparrow}$, $M_{t^-}^{(i)} = \psi_l(1 - r_l)M_{t^+}^{(i-1)}$

2. removed sampled leaf:

    (a) in $\mathcal{T}_t^{\downarrow}$, $L_{t^+}^{(i)} = \psi_l r_l L_{t^-}^{(i)}$

    (b) in $\mathcal{T}_t^{\uparrow}$, $M_{t^-}^{(i)} = \psi_l r_l M_{t^+}^{(i)}$

3. sampling along a branch:

    (a) in $\mathcal{T}_t^{\downarrow}$, $L_{t^+}^{(i)} = \psi_l(1 - r_l)L_{t^-}^{(i)}$

    (b) in $\mathcal{T}_t^{\uparrow}$, $M_{t^-}^{(i)} = \psi_l(1 - r_l)M_{t^+}^{(i)}$

4. occurrence:

    (a) in $\mathcal{O}_t^{\downarrow}$, $L_{t^+}^{(i)} = (k + i)\omega_l(1 - r_l)L_{t^-}^{(i)}$

    (b) in $\mathcal{O}_t^{\uparrow}$, $M_{t^-}^{(i)} = (k + i)\omega_l(1 - r_l)M_{t^+}^{(i)}$

5. removed occurrence:

    (a) in $\mathcal{O}_t^{\downarrow}$, $L_{t+}^{(i)} = \omega_l r_l i L_{t-}^{(i-1)}$

    (b) in $\mathcal{O}_t^{\uparrow}$, $M_{t-}^{(i)} = \omega_l r_l (i+1) M_{t+}^{(i+1)}$

6. branching event:

    (a) in $\mathcal{T}_t^{\downarrow}$, $L_{t+}^{(i)} = \lambda_l L_{t-}^{(i)}$

    (b) in $\mathcal{T}_t^{\uparrow}$, $M_{t-}^{(i)} = \lambda_l M_{t+}^{(i)}$

Numerical approximation of the ODEs

As described above, for any constant rate time interval where $\tau_l \leqslant t < \tau_{l+1}$, M_t and L_t are defined along epochs as the solution to systems of differential equations S8 and S10 for $t_h \leqslant t < t_{h+1}$. Numerically, the solution to such systems of equations is approximated by truncating the system at a fixed integer N as follows:

$$L_{t_{h+1}} = e^{A_l(t-t_h)} L_{t_h} \tag{S11}$$

$$M_{t_h} = e^{A_l'(t-t_{h+1})} M_{t_{h+1}} \tag{S12}$$

Where A_l and A_l' are $N \times N$ tridiagonal matrices with ODE coefficients. When there is a rate shift τ_l within an epoch (t_h, t_{h+1}), the epoch is cut in two parts and L_t and M_t are simply computed as,

$$L_{t_{h+1}} = e^{A_{l+1}(t_{h+1}-\tau_l)} e^{A_l(\tau_l-t_h)} L_{t_h} \tag{S13}$$

$$M_{t_h} = e^{A_l'(t_h-\tau_l)} e^{A_{l+1}'(\tau_l-t_{h+1})} M_{t_{h+1}} \tag{S14}$$

This can be extended to any number of rate changes within an epoch. This strategy of solving for L_t and M_t yields the following two algorithms. Because exponential matrices are computationally intensive to calculate, these algorithms are only used in the most general cases, when no other analytical formula is available (i.e. when $\omega \neq 0$ and $r \neq 1$).

Algorithm 1 Computes a numerical approximation of L_t for a specific set of times with known rate shift events

---

Input:

Observed tree and occurrence data $(\mathcal{T}, \mathcal{O})$,

extant sampling probability ρ,

set of times of rate shift events $(\tau_l)_{l=0}^{m+1}$,

and corresponding sets of parameters :

vector $\lambda = (\lambda_l)_{l=0}^{m}$ where λ_l is the birth rate in time interval $[\tau_l, \tau_{l+1})$

vector $\mu = (\mu_l)_{l=0}^{m}$ where μ_l is the death rate in time interval $[\tau_l, \tau_{l+1})$

vector $\psi = (\psi_l)_{l=0}^{m}$ where ψ_l is the sampling rate in time interval $[\tau_l, \tau_{l+1})$

vector $\omega = (\omega_l)_{l=0}^{m}$ where ω_l is the rate of occurence sampling in time interval $[\tau_l, \tau_{l+1})$

vector $r = (r_l)_{l=0}^{m}$ where r_l is the removal probability in time interval $[\tau_l, \tau_{l+1})$

set of time points $(d_j)_{j=1}^{S}$ for which we want to compute the density, and

the truncation N setting the accuracy of the algorithm.

Output: A numerical approximation of L_t at times $(d_j)_{j=1}^{S}$, $(\widetilde{L}_t^{(i)})_{\substack{i \in \{0,1,\ldots,N\} \\ j \in \{1,2,\ldots,S\}}}$.

1: Pool all $(d_j)_{j=1}^{S}$, all branching and sampling times of $(\mathcal{T}, \mathcal{O})$ and rate shift times $(\tau_l)_{l=0}^{m+1}$ in an ordered list $(t_h)_{h=1}^{n+m+1}$.

2: Set $j = 1$ and initialize B as a $S \times N + 1$ empty matrix.

3: Set $l = 0$ and $\lambda = \lambda_0$, $\mu = \mu_0$, $\psi = \psi_0$, $\omega = \omega_0$, $r = r_0$, $\gamma_0 = \lambda_0 + \mu_0 + \psi_0 + \omega_0$.

4: Set $\forall i \in \{0, 1, \ldots, N\}$, $\widetilde{L}_0^{(i)} = \rho^{k_0}(1-\rho)^i$.

5: **for** $h = 1, 2, \ldots, n + m + 1$ **do**

6:     Numerically solve the ODE $\dot{\widetilde{L}}_t = A\widetilde{L}_t$ on (t_h, t_{h+1}), where matrix A is a $N \times N$ tridiagonal matrix with entries given by,

$$\forall i \in \{0, 1, \ldots, N\} \quad A^{(i,i)} = \gamma(k+i)$$

$$\forall i \in \{0, 1, \ldots, N-1\} \quad A^{(i,i+1)} = \lambda(2k+i)$$

$$\forall i \in \{1, 2, \ldots, N\} \quad A^{(i,i-1)} = \mu i$$

7:     **if** $t_h = d_j$ **then**

8:         Set $B^{(j,i)} = \widetilde{L}_{t_h}^{(i)}$ and

9:         Set $j = j + 1$.

10:    **end if**

Algorithm 2 Computes a numerical approximation of M_t for a specific set of times with known rate shift events

---

Input:

Observed tree and occurrence data $(\mathcal{T}, \mathcal{O})$,

parameters t_{or}, ρ

set of times of rate shift events $(\tau_l)_{l=0}^{m+1}$,

and corresponding sets of parameters :

vector $\lambda = (\lambda_l)_{l=0}^{m}$ where λ_l is the birth rate in time interval $[\tau_l, \tau_{l+1})$

vector $\mu = (\mu_l)_{l=0}^{m}$ where μ_l is the death rate in time interval $[\tau_l, \tau_{l+1})$

vector $\psi = (\psi_l)_{l=0}^{m}$ where ψ_l is the sampling rate in time interval $[\tau_l, \tau_{l+1})$

vector $\omega = (\omega_l)_{l=0}^{m}$ where ω_l is the rate of occurence sampling in time interval $[\tau_l, \tau_{l+1})$

vector $r = (r_l)_{l=0}^{m}$ where r_l is the removal rate in time interval $[\tau_l, \tau_{l+1})$

set of time points $(d_j)_{j=1}^{S}$ for which we want to compute the density,

and the truncation N setting the accuracy of the algorithm.

Output: A numerical approximation of M_t at times $(d_j)_{j=1}^{S}$, $(\widetilde{M}_t^{(i)})_{\substack{i \in \{0,1,\ldots,N-1\} \\ j \in \{1,2,\ldots,S\}}}$.

1: Pool all (d_j), rate shift times (τ_l) and all branching and sampling times of $(\mathcal{T}, \mathcal{O})$ in an ordered list $(t_h)_{h=1}^{n}$.

2: Set $j = S$, $k = m$ and B' as a $S \times N$ empty matrix.

3: Set $\forall i \in \{0, 1, \ldots, N-1\}$, $\widetilde{M}_{t_n}^{(i)} = \mathbb{1}_{i=0}$.

4: Set $l = m$ and $\lambda = \lambda_m$, $\mu = \mu_m$, $\psi = \psi_m$, $\omega = \omega_m$, $r = r_m$.

5: **for** $h = n-1, n-2, \ldots, 0$ **do**

6:     Numerically solve the ODE $\dot{\widetilde{M}}_t = A' \widetilde{M}_t$ on (t_h, t_{h+1}), where matrix A' is a $N \times N$ tridiagonal matrix with entries given by,

$$\forall i \in \{0, 1, \ldots, N-1\} \quad A'^{(i,i)} = \gamma(k+i)$$

$$\forall i \in \{0, 1, \ldots, N-2\} \quad A'^{(i,i+1)} = -\mu(i+1)$$

$$\forall i \in \{1, 2, \ldots, N-1\} \quad A'^{(i,i-1)} = -\lambda(2k+i-1)$$

7:     **if** $t_h = \tau_j$ **then**

8:         Set $B'^{(j,i)} = \widetilde{M}_{t_h}^{(i)}$ and $j = j - 1$.

9:     **end if**

10:     **if** $t_h = 0$ or $t_h = \tau_S$ **then**

B – Extension of analytical results

Here, we aim at extending some analytical results of Gupta et al. (2020) and Manceau et al. (2021) to a piecewise-constant parameter setting. We start with the probability of extinction before time t of a process starting at 0 with one lineage, u_t. We then detail p_t, the probability that a lineage starting at time 0 leads to one sampled lineage at time t. Finally, we detail what happens to L_t and M_t for specific subcases, when $\omega = 0$ or $r = 1$. Note that formulas for u and p with rate shifts can be found in Stadler et al. (2013) as well.

The extinction probability across rate shifts

Let's start slowly with u, one time slice after the other.

On the first time slice We start with some initializing condition, say, $u_0 = z$.

Then, on $(\tau_0 = 0, \tau_1)$, we have a first set of parameters $(\lambda_0, \mu_0, \gamma_0)$ and u satisfies the following ODE,

$$\dot{u}_s = \lambda_0 u_s^2 - \gamma_0 u_s + \mu_0$$

which solution can be written as,

$$\forall t \in (\tau_0, \tau_1), \quad u_t = \frac{x_0^{(1)}(x_0^{(2)} - z) - x_0^{(2)}(x_0^{(1)} - z)e^{-\sqrt{\Delta}t}}{(x_0^{(2)} - z) - (x_0^{(1)} - z)e^{-\sqrt{\Delta}t}}$$

where $\Delta_0 = \gamma_0^2 - 4\lambda_0\mu_0$ and $x_0^{(1)}$ and $x_0^{(2)}$ are the roots of the polynomial $\lambda_0 x^2 - \gamma_0 x + \mu_0$, i.e.,

$$x_0^{(1)} = \frac{\gamma_0 - \sqrt{\Delta_0}}{2\lambda_0} \quad \text{and} \quad x_0^{(2)} = \frac{\gamma_0 + \sqrt{\Delta_0}}{2\lambda_0}$$

At the end of the time slice, we thus get,

$$u_{\tau_1} = \frac{x_0^{(1)}(x_0^{(2)} - z) - x_0^{(2)}(x_0^{(1)} - z)e^{-\sqrt{\Delta_0}\tau_1}}{(x_0^{(2)} - z) - (x_0^{(1)} - z)e^{-\sqrt{\Delta_0}\tau_1}}$$

On the second time slice We now start with initial condition u_{τ_1}.

Then, on (τ_1, τ_2), we have a second set of parameters $(\lambda_1, \mu_1, \gamma_1)$ and u satisfies the following ODE with these new parameters:

$$\dot{u}_s = \lambda_1 u_s^2 - \gamma_1 u_s + \mu_1$$

which solution can be written as,

$$\forall t \in (\tau_1, \tau_2), \quad u_t = \frac{x_1^{(1)}(x_1^{(2)} - u_{\tau_1}) - x_1^{(2)}(x_1^{(1)} - u_{\tau_1})e^{-\sqrt{\Delta_1}(t-\tau_1)}}{(x_1^{(2)} - u_{\tau_1}) - (x_1^{(1)} - u_{\tau_1})e^{-\sqrt{\Delta_1}(t-\tau_1)}}$$

And so on and so forth In doing so, we get that computing u_t for a given time t thus requires recursively computing u_0, and then u_{τ_1}, u_{τ_2}, ... until getting to u_{τ_l}, where $\tau_l \leqslant t \leqslant \tau_{l+1}$.

$$\forall t \in (\tau_l, \tau_{l+1}), \quad u_t = \frac{x_l^{(1)}(x_l^{(2)} - u_{\tau_l}) - x_l^{(2)}(x_l^{(1)} - u_{\tau_l})e^{-\sqrt{\Delta_l}(t-\tau_l)}}{(x_l^{(2)} - u_{\tau_l}) - (x_l^{(1)} - u_{\tau_l})e^{-\sqrt{\Delta_l}(t-\tau_l)}}$$

The probability to see one lineage across rate shifts

Let's apply carefully the same method now for p.

On the first time slice We start with some initializing condition $p_0 = 1 - z$.

Then on (τ_0, τ_1), we have a first set of parameters and p satisfies,

$$\dot{p}_s = (2\lambda_0 u_s - \gamma_0)p_s$$

which solution at first is the same as without skyline changes, i.e.

$$p_t = (1 - z)\frac{\Delta_0}{\lambda_0^2}\left((x_0^{(2)} - z) - (x_0^{(1)} - z)e^{-\sqrt{\Delta_0}t}\right)^{-2} e^{-\sqrt{\Delta_0}t}$$

On the second time slice We start now with some initializing condition p_{τ_1} and would like to solve the following ODE on (τ_1, τ_2),

$$\dot{p}_s = (2\lambda_1 u_s - \gamma_1)p_s$$

Replacing the expression of u_s on this time slice gives us,

$$\frac{dp_s}{p_s} = \left(2\lambda_1 \frac{x_1^{(1)}(x_1^{(2)} - u_{\tau_1}) - x_1^{(2)}(x_1^{(1)} - u_{\tau_1})e^{-\sqrt{\Delta_1}(s-\tau_1)}}{(x_1^{(2)} - u_{\tau_1}) - (x_1^{(1)} - u_{\tau_1})e^{-\sqrt{\Delta_1}(s-\tau_1)}} - \gamma_1\right) ds$$

We thus end up with

$$\forall t \in (\tau_1, \tau_2), \quad p_t = p_{\tau_1} \frac{\Delta_1}{\lambda_1^2} \left((x_1^{(2)} - u_{\tau_1}) - (x_1^{(1)} - u_{\tau_1})e^{-\sqrt{\Delta_1}(t-\tau_1)} \right)^{-2} e^{-\sqrt{\Delta_1}(t-\tau_1)}$$

And so on and so forth This gives us

$$\forall t \in (\tau_l, \tau_{l+1}), \quad p_t = p_{\tau_l} \frac{\Delta_l}{\lambda_l^2} \left((x_l^{(2)} - u_{\tau_l}) - (x_l^{(1)} - u_{\tau_l})e^{-\sqrt{\Delta_l}(t-\tau_l)} \right)^{-2} e^{-\sqrt{\Delta_l}(t-\tau_l)}$$

Using these for computation of L without occurrences

When $\omega = 0$, we can still use the ansatz $L_t^{(i)} = u_t^i W_t$ and look for W_t. On a given epoch, the ODE on $L_t^{(i)}$ translates as $\dot{W}_t = (2\lambda u_t - \gamma)kW_t$.

Solving this between time t and t_h, on time slice number l, leads us to

$$\begin{aligned} W_t &= W_{t_h} \left(\frac{(x_l^{(2)} - u_{\tau_l}) - (x_l^{(1)} - u_{\tau_l})e^{-\sqrt{\Delta_l}(t-\tau_l)}}{(x_l^{(2)} - u_{\tau_l}) - (x_l^{(1)} - u_{\tau_l})e^{-\sqrt{\Delta_l}(t_h-\tau_l)}} \right)^{-2k} e^{-k\sqrt{\Delta_l}(t-t_h)} \\ &= W_{t_h} \left(\frac{p(t)}{p(t_h)} \right)^k \end{aligned}$$

With this last equality still holding true, the induction across all epochs remains identical to the what was described in Manceau et al. (2021).

Using these for the computation of M without occurrences

What happens to the PDE solution over successive time slices with different parameters, when $\omega = 0$? Let's start slowly again, one time slice after the other.

On the first time slice We assume here that (t_{h-1}, t_h) is an epoch with $t_h \leqslant \tau_1$, such that we are still in the first time slice with parameters $(\lambda_0, \mu_0, \gamma_0)$. The PDE is

$$\begin{aligned} \widehat{M}(t_h, z) &= F(z) \\ \partial_t \widehat{M} + (\lambda_0 z^2 - \gamma_0 z + \mu_0)\partial_z \widehat{M} + k(2\lambda_0 - \gamma_0)\widehat{M} &= 0 \end{aligned}$$

We use the method of characteristics as for the constant-parameter case, writing

137    $g(s) = \widehat{M}(t(s), z(s))$ with functions t, z and g satisfying

$$\frac{dt}{ds} = 1$$

$$\frac{dz}{ds} = \lambda_0 z^2 - \gamma_0 z + \mu_0$$

$$\frac{dg}{ds} = -k(2\lambda_0 z - \gamma_0)g$$

138   We thus keep $t(s) = t_h + s$, i.e. $s = t - t_h$.

139   Then, turning to $z(s)$, we get

$$z(s) = u_0(s, z_0) = \frac{x_0^{(1)}(x_0^{(2)} - z_0) - x_0^{(2)}(x_0^{(1)} - z_0)e^{-\sqrt{\Delta_0}s}}{(x_0^{(2)} - z_0) - (x_0^{(1)} - z_0)e^{-\sqrt{\Delta_0}s}}$$

140   thus leading to $z_0 = u_0(t_h - t, z)$, where u_0 denotes the above explicitly defined function.

141   Note that on this time slice, $\forall t, \quad t_h - t \leqslant \tau_1$, so $u_0 = u$ here. But on successive time slices

142   it will not be the case anymore.

143   Finally, we get, for g, the following,

$$g_s = g_0 \left(\frac{(x_0^{(2)} - z_0) - (x_0^{(1)} - z_0)e^{-\sqrt{\Delta_0}s}}{x_0^{(2)} - x_0^{(1)}} \right)^{2k} e^{k\sqrt{\Delta_0}s}$$

$$= g_0 \left(\frac{1 - z}{p_0(s, z_0)} \right)^k$$

144   where we denote here again by p_0 the function p as in the constant-parameter case with

145   parameters $(\lambda_0, \mu_0, \gamma_0)$.

146   As a result, we get

$$\widehat{M}(t, z) = F(u_0(t_h - t, z))R_0(t_h - t, z)^k$$

147   *And so on and so forth* Because nothing really simplifies at this stage, we get the same on

148   following time slices. On time slice l, we only change the indices and consider functions u

149   and R as in the constant-parameter case with parameters $(\lambda_l, \mu_l, \gamma_l)$,

$$\widehat{M}(t, z) = F(u_l(t_h - t, z))R_l(t_h - t, z)^k$$

150   *What happens to the induction across epochs* We thus hope that simplifications will

151   appear in the induction across epochs. In order to make them appear, we'll define here

functions of three variables instead of only two. We keep the same names, so I hope it'll not be too confusing.

Starting now, we introduce a function of three variables u, where value $u(t_1, t_0, z)$ is the probability that one lineage starting at time t_1 in the past, goes extinct/unsampled before time t_0, knowing there is a field of bullets with intensity z at time t_0. On a single time slice, this is the solution of the usual ODE driving the evolution of u, but with initial condition $u_{t_0} = z$ instead of $u_0 = z$.

Let's then do the same with function p, defining $p(t_1, t_0, z)$ as the probability that one lineage starting at time t_1 in the past leads to one sampled lineage at time t_0, knowing there is a field of bullets of intensity z at time t_0. On a single time slice, this is the solution of the usual ODE driving the evolution of p, but with initial condition $p_{t_0} = 1 - z$.

Note now that across time slices, if $t_2 \geqslant \tau_l$ and $t_0 \leqslant \tau_l$, then $u(t_2, t_0, z)$ can be computed as the solution of the usual ODE with parameters $(\lambda_{l-1}, \gamma_{l-}, \mu_{l-1})$, with initial condition $u_{t_0} = z$, until getting $u(\tau_l, t_0, z)$. Then the ODE with parameter set $(\lambda_l, \gamma_l, \mu_l)$ is used, with initial value $u(\tau_l, t_0, z)$, until getting $u(t_2, t_0, z)$. More explicitly, this gives us,

$$u(\tau_l, t_0, z) = \frac{x_{l-1}^{(1)}(x_{l-1}^{(2)} - z) - x_{l-1}^{(2)}(x_{l-1}^{(1)} - z)e^{-\sqrt{\Delta_{l-1}}(\tau_l - t_0)}}{(x_{l-1}^{(2)} - z) - (x_{l-1}^{(1)} - z)e^{-\sqrt{\Delta_{l-1}}(\tau_l - t_0)}}$$

$$u(t_2, t_0, z) = \frac{x_l^{(1)}(x_l^{(2)} - u(\tau_l, t_0, z)) - x_l^{(2)}(x_l^{(1)} - u(\tau_l, t_0, z))e^{-\sqrt{\Delta_l}(t_2 - \tau_l)}}{(x_l^{(2)} - u(\tau_l, t_0, z)) - (x_l^{(1)} - u(\tau_l, t_0, z))e^{-\sqrt{\Delta_l}(t_2 - \tau_l)}}$$

To recursively compute $p(t_2, t_0, z)$ across time slices, we would need,

$$p(\tau_l, t_0, z) = (1 - z)\left(\frac{(x_{l-1}^{(2)} - z) - (x_{l-1}^{(1)} - z)e^{\sqrt{\Delta_{l-1}}\tau_l}}{(x_{l-1}^{(2)} - z) - (x_{l-1}^{(1)} - z)e^{\sqrt{\Delta_{l-1}}t_0}}\right)^{-2} e^{-\sqrt{\Delta_{l-1}}(\tau_l - t_0)}$$

$$p(t_2, t_0, z) = p(\tau_l, t_0, z)\left(\frac{(x_l^{(2)} - u(\tau_l, t_0, z)) - (x_l^{(1)} - u(\tau_l, t_0, z))e^{\sqrt{\Delta_l}t_2}}{(x_l^{(2)} - u(\tau_l, t_0, z)) - (x_l^{(1)} - u(\tau_l, t_0, z))e^{\sqrt{\Delta_l}\tau_l}}\right)^{-2} e^{-\sqrt{\Delta_l}(t_2 - \tau_l)}$$

We are now especially interested in the property that is at the core of the induction, i.e. formerly,

$$R(t_{or} - t_h, u(t_h - t, z)) = \frac{R(t_{or} - t, z)}{R(t_h - t, z)}$$

which we would like to extend as,

$$R(t_{or}, t_h, u(t_h, t, z)) = \frac{R(t_{or}, t, z)}{R(t_h, t, z)} \tag{S15}$$

We first need to show that

$$u(t_2, t_1, u(t_1, t_0, z)) = u(t_2, t_0, z)$$
$$\frac{p(t_2, t_1, u(t_1, t_0, z))}{1 - u(t_1, t_0, z)} = \frac{p(t_2, t_0, z)}{p(t_1, t_0, z)}$$

The first equation seems quite natural, thanks to the semi-group property of solutions of ODEs (or thanks to the probabilistic interpretation of these quantities). For the second one, we can check by calculus that it is correct whether (t_0, t_1, t_2) are in the same time slice or not.

$$\frac{p(t_2, t_1, u(t_1, t_0, z))}{1 - u(t_1, t_0, z)} = \frac{1 - u(t_1, t_0, z)}{1 - u(t_1, t_0, z)} \left(\frac{(x_l^{(2)} - u(t_1, t_0, z)) - (x_l^{(1)} - u(t_1, t_0, z))e^{-\sqrt{\Delta_l}t_2}}{(x_l^{(2)} - u(t_1, t_0, z)) - (x_l^{(1)} - u(t_1, t_0, z))e^{-\sqrt{\Delta_l}t_1}} \right)^{-2} e^{-\sqrt{\Delta_l}(t_2 - t_1)}$$

$$= \frac{p(t_1, t_0, z)}{p(t_1, t_0, z)} \left(\frac{(x_l^{(2)} - u(t_1, t_0, z)) - (x_l^{(1)} - u(t_1, t_0, z))e^{-\sqrt{\Delta_l}t_2}}{(x_l^{(2)} - u(t_1, t_0, z)) - (x_l^{(1)} - u(t_1, t_0, z))e^{-\sqrt{\Delta_l}t_1}} \right)^{-2} e^{-\sqrt{\Delta_l}(t_2 - t_1)}$$

$$= \frac{p(t_2, t_0, z)}{p(t_1, t_0, z)}$$

This property on p thus ensures the equality (S15), which in turn allows us to carry out our induction across epochs in the skyline version as

$$\widehat{M}(t, z) = \lambda^x \psi^{v+w+y} r^w (1-r)^{v+y} \prod_{t_j \in \mathcal{X} \cup \{t_{or}\}} R(t_j, t, z) \prod_{t_j \in \mathcal{W}} R(t_j, t, z)^{-1} \prod_{t_j \in \mathcal{Y}} u(tj, t, z)(R(t_j, t, z)^{-1}$$

$$\tag{S16}$$

C − RevBayes implementation

Core algorithms

To enable great flexibility and ensure fast computation, RevBayes is constructed

around a mirror structure (Fig. S5) in which all the core functions coded in C++ are

reflected in the revlanguage section that links with the Rev language interface.
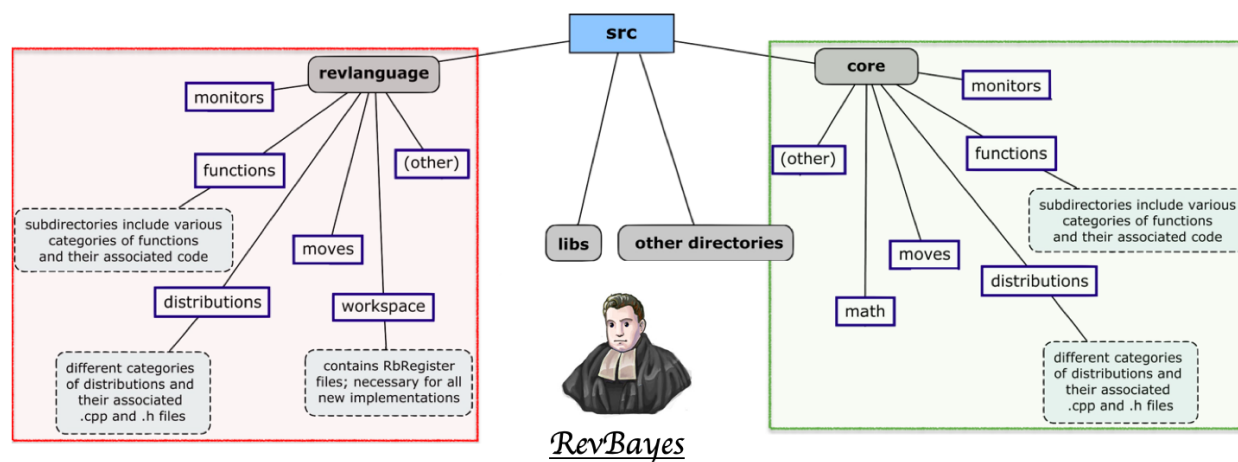
Figure S5. Simplified representation of the RevBayes structure. Modified from the RevBayes website, keeping only descriptions of the folders we modified. Note the organizational symmetry between the core directory containing the hard-coded features and the revlanguage directory matching the Rev syntax.

Due the multiple advantages of RevBayes and its increasing use, particularly for

macroevolutionary research, we chose this software to implement the OBDP. All our

modifications have been carried out in a separate copy of its development branch on

GitHub (https://github.com/revbayes/revbayes/tree/dev-cevo-lab), and are aimed

to be integrated in a future stable release. They consist in 3 key additions detailed in Table

S1.

The necessary first step was to implement the core algorithms responsible for

computing the quantities L_t and M_t through time. The final organisation is as follows:

from outside of the *ComputeLikelihoodsLtMt.cpp* file (see Table S1) the only functions

called are *ComputeLnProbabilityDensitiesOBDP* – returning L_t and M_t through time – or

ComputeLnLikelihoodOBDP – returning only the final likelihood. Those functions will

¹⁹⁴ themselves call the appropriate internal function (*ForwardsTraversalMt* or

¹⁹⁵ *BackwardsTraversalLt*) with the correct parameters. Those rely on a key function,

¹⁹⁶ *PoolEvents*, the role of which is to construct the vector containing all the events that will

¹⁹⁷ be browsed by the traversal algorithms, namely branching times, ψ- and ω-sampling times,

¹⁹⁸ and time points for which we want to store the probability distribution.

¹⁹⁹ Because the densities computed during the traversals very quickly reached

²⁰⁰ excessively small or elevated values, to the point of exceeding the maximum number of

²⁰¹ recorded decimals, a correction term is added at each step to bring the densities closer to

²⁰² 1. At the end of the traversal, the recorded correction terms plus the factorizable factors

²⁰³ are added to the log-transformed densities.
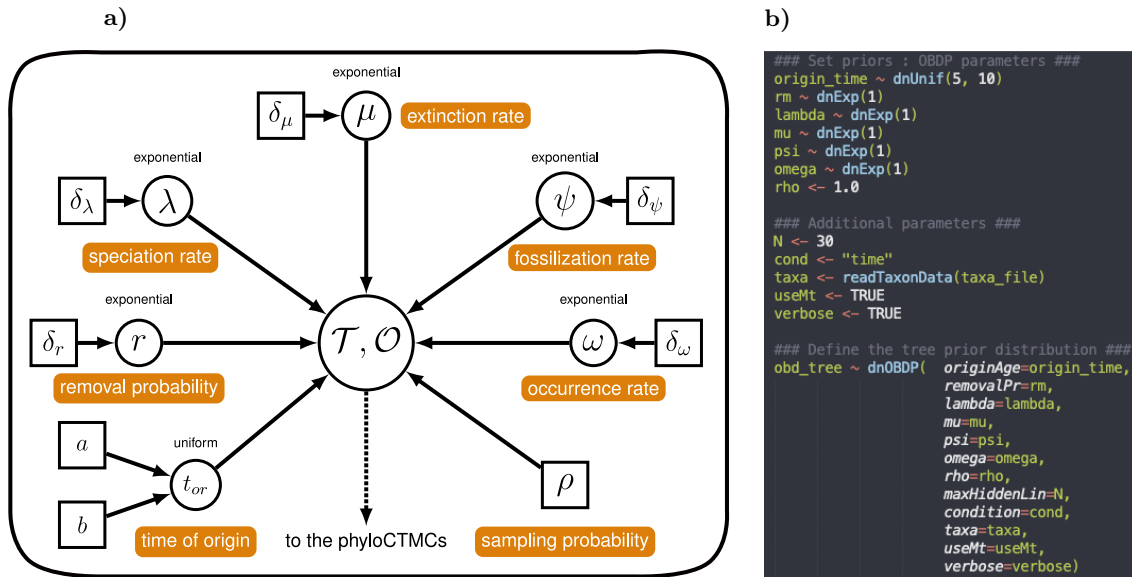
Figure S6. A graphical model of the OBDP and its translation into the Rev language. a) Graphical model, modified from the RevBayes FBD tutorial, representing the OBDP parameters – labelled in orange – generating a reconstructed tree $\mathcal{T}$ and a record of occurrences $\mathcal{O}$. b) Rev script corresponding to this graphical model. Note the distinction between the $\sim$ notation attributing a distribution to a stochastic node and the $\leftarrow$ notation defining a constant node.

²⁰⁴ In addition, the Occurrence Birth-Death Process and the traversal algorithms not

²⁰⁵ only allow us to perform a MCMC phylogenetic inference incorporating the occurrences,

²⁰⁶ they can also be used to output the probability distribution of the number of lineages

Table S1. Overview of the implementations carried out to incorporate the Occurrence Birth-Death Process and the associated Diversity Inference method into RevBayes. It lists for each of our goals the associated C++ files, along with their assignment in the RevBayes structure.

Objectives	Location	File names	Major new functions
1. Perform Forwards and Backwards traversal algorithms	core/ functions	*ComputeLikelihoods LtMt.h ComputeLikelihoods LtMt.cpp*	*ComputeLnProbability- DensitiesOBDP ComputeLnLikelihoodOBDP PoolEvents ForwardsTraversalMt BackwardsTraversalLt*
2. Encode the OBDP distribution	core/ distributions	*OccurrenceBirthDeath Process.h OccurrenceBirthDeath Process.cpp*	*OccurrenceBirthDeathProcess computeLnProbability- DivergenceTimes*
	revlanguage/ distributions	*Dist_occurrenceBirth DeathProcess.cpp Dist_occurrenceBirth DeathProcess.h*	*createDistribution getParameterRules*
3. Infer past diversity	core/ distributions	*InferAncestralPop SizeFunction.h InferAncestralPop SizeFunction.cpp*	*InferAncestralPopSizeFunction*
	revlanguage/ distributions	*Func_inferAncestral PopSize.h Func_inferAncestral PopSize.cpp*	*createFunction getArgumentRules*

207 through time, K_t. We introduced this functionality into RevBayes through

208 *InferAncestralPopSizeFunction*, which can be called directly from the Rev interface. As

209 with the OBDP distribution, we had to design the parameter loading procedure, then call

210 the *ComputeLnProbabilityDensitiesOBDP* function to get the $\log(L_t)$ and $\log(M_t))$

211 matrices and finally combine and normalize them to obtain the $\log(K_t))$ matrix.

212   *RevGadgets*

213   The postprocessing step consists in computing the posterior probability of the total

214 number of lineages through time. It can be performed independently of the previous steps,

given that one has at least a tree, a set of parameters and optionally occurrence times. It comprises 2 steps, the first one uses the *fnInferAncestralPopSize* function, implemented in RevBayes, to obtain the matrix of diversity densities K_t for each tree in the MCMC trace. Then, in order to convert K_t matrices into a nicely rendered plot we added two functions in the auxiliary R library RevGadgets (Tribble et al., 2021). Starting from the trace of posterior trees, parameters, and K_t matrices one first needs to execute the *rev.process.nbLineages* function that will organize the required information into the *Kt_mean* data frame. The goal is to incorporate all the uncertainty concerning the inferred parameter values and tree topologies into the diversity trajectory estimation. Afterwards, this averaged *Kt_mean* is used by the function *rev.plot.nbLineages* to realize the final plot using ggplot2 (Wickham, 2016). Here it is possible to alter most of the display options, such as the types of lineages to be shown (observed, hidden, total), as well as their colours and shapes (see e.g. Fig. S8).

Table S2. Description of two novel RevGadgets functions for visualizing OBDP diversity-through-time estimations. The input objects and display parameters are detailed, those with an asterisk always have to be provided while the others have default values.

Function	Option	Type	Description
rev.process .nbLineages	*start_time_trace_file**	*character*	*MCMC trace of the starting times.*
	popSize_distribution _matrices_file*	*character*	*Matrices computed with fnInferAncestralPopSize in RevBayes.*
	trees_trace_file*	*character*	*MCMC trace of the trees.*
	weight_trees_posterior	*Boolean*	*Whether to combine trees uniformly or weighted by their posterior probabilities.*
rev.plot .nbLineages	*Kt_mean**	*data.frame*	*Processed output for plotting.*
	xlab / ylab	*character*	*Label of the x-axis / y-axis.*
	line.size / interval.line.size	*numeric*	*Width of the lineage plot / credible interval line.*
	col.Hidden / col.Observed / col.Total / col.Hidden.interval / col.Total.interval	*character*	*Color of the hidden / observed / total lineages plot line. Color of the credible interval for hidden / total lineages.*
	palette.Hidden / palette.Total	*character*	*Palette of the hidden / total lineages distribution.*
	show.Hidden / show.Observed / show.Total / show.intervals / show.densities / show.expectations	*Boolean*	*Whether to show the plot for hidden / observed / total lineages / credible intervals / diversity densities / diversity expectations.*
	use.interpolate	*Boolean*	*Whether to interpolate densities.*

₂₂₈   D − Qualitative validation: "blind test" on simulated data

₂₂₉   Parameter values used to simulate the two datasets used in the blind test are

₂₃₀   presented in Table S3. Two trees with occurrences have been simulated under the OBDP

₂₃₁   (parameters 1-6). For "dataset 1", genetic sequences along the first tree are simulated

₂₃₂   according to a K80 model of molecular evolution (parameters 7-9) and recorded only for

₂₃₃   extant taxa. Binary traits are simulated according to a Markov process with symmetrical

₂₃₄   rates (parameters 10-12) and are recorded for both extant and extinct taxa. This

₂₃₅   corresponds to a classic macroevolution scenario. For "dataset 2", genetic sequences along

₂₃₆   the second tree are simulated according to a K80 model of molecular evolution (parameters

₂₃₇   7-9) and recorded for extant and extinct individuals. This allows us to have a better

₂₃₈   resolution of the underlying tree than in the first dataset. Moreover, getting genetic

₂₃₉   sequences for individuals sampled in the past corresponds more to an epidemiology

₂₄₀   scenario.

Table S3. Parameter values used to simulate two datasets and test our OBDP inference workflow.

λ	μ	ψ	ω	r	ρ	m_{nt}	α_{nt}	β_{nt}	m_{morpho}	q_{01}	q_{10}
1	0.9	0.2	0.3	0	0.8	10000	0.01	0.02	60	0.03	0.03

₂₄₁   Two of us, ignorant of the values used for simulation, designed the inference

₂₄₂   protocol and conducted the analysis, taking as input the occurrences, sequences, and

₂₄₃   morphological data only. Priors used for inference on "dataset 1" are presented in Table S4

₂₄₄   and the general setup for analysis is illustrated in Figure S7. Priors used for inference on

₂₄₅   "dataset 2" were very similar, except for the absence of a model of morphological

₂₄₆   evolution, and they are presented in Table S5.
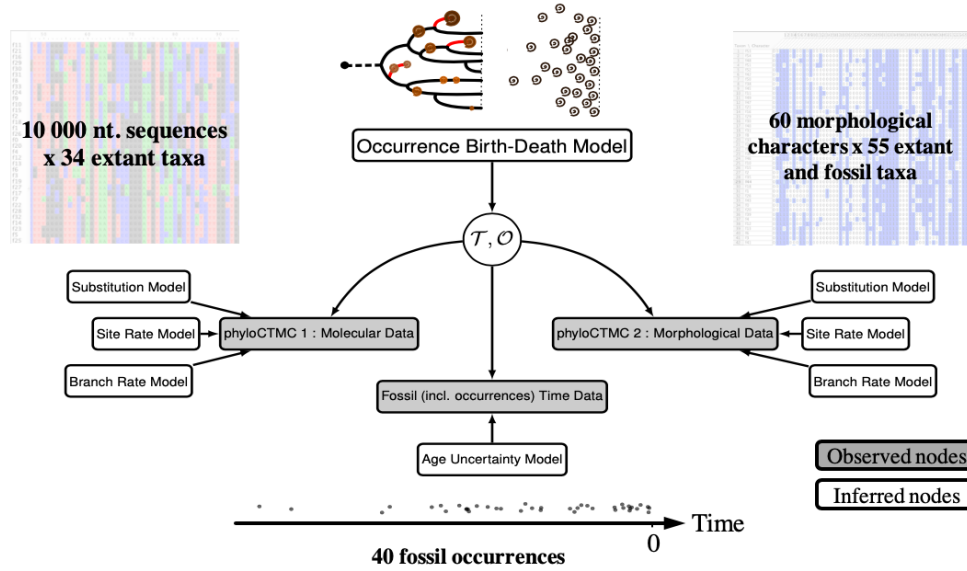
Figure S7. Modular representation of the graphical models used in the qualitative validation analysis. Modified from [Heath et al. (2019)](#). The simulated data, noted in the grey nodes are used to deduce the posterior distributions of all other random variables noted in the white nodes.

Table S4. Prior distributions on the OBDP parameters and models for the "Blind Test" analysis on dataset 1. Notations: $\mathcal{U}$ for the Uniform distribution, $\mathcal{E}$ for Exponential, Dir for Dirichlet, GTR for the General Time Reversible substitution model and MK for the Mk model, the analog of JC69 for an arbitrary number of character states.

Parameter	Prior	Model	Prior
λ	$\mathcal{E}(10)$		Strict clock rate: $\mathcal{E}(10)$
μ	$\mathcal{E}(10)$	*Molecular evolution:*	Exchangeability rates: $Dir(1,1,1,1,1,1)$
ψ	$\mathcal{E}(10)$	$GTR + \Gamma$	Stationary frequencies: $Dir(1,1,1,1)$
ω	$\mathcal{E}(5)$		Gamma distribution shape: $\mathcal{E}(1)$
ρ	$\mathcal{U}(0,1)$	Morphological evolution:	Strict clock rate: $\mathcal{E}(1)$
r	0	$MK + \Gamma$	Gamma distribution shape: $\mathcal{E}(1)$
t_{or}	$\mathcal{U}(7.7,12)$		

Table S5. Prior distributions of the OBDP parameters and models for the "Blind Test" analysis on dataset 2. Notations: $\mathcal{U}$ for the Uniform distribution, B for the Beta distribution, $\mathcal{E}$ for Exponential, Dir for Dirichlet, GTR for the General Time Reversible substitution model.

Parameter	Prior	Model	Prior
λ	$\mathcal{E}(10)$		Strict clock rate: $\mathcal{E}(10)$
μ	$\mathcal{E}(10)$	*Molecular evolution:*	Exchangeability rates: $Dir(1,1,1,1,1,1)$
ψ	$\mathcal{E}(10)$	$GTR + \Gamma$	Stationary frequencies: $Dir(1,1,1,1)$
ω	$\mathcal{E}(10)$		Gamma distribution shape: $\mathcal{E}(1)$
ρ	$B(1.0,1.0)$		
r	0		
t_{or}	$\mathcal{U}(7.7,12)$		

²⁴⁷   In our blind inferences, we recovered posterior distribution of diversity trajectories

²⁴⁸ (Fig. S8) and trees (Fig. S9) which are very close to the real data from the simulations.

²⁴⁹ The true number of hidden lineages is most of the time near the expectation of the inferred

²⁵⁰ posterior distribution and more importantly always in the 95% posterior credible interval.

²⁵¹ When looking at the total number of lineages – i.e. species richness in macroevolution or

²⁵² prevalence in epidemiology – the estimates remains very close to the truth and almost

²⁵³ always in the 95% credible interval.
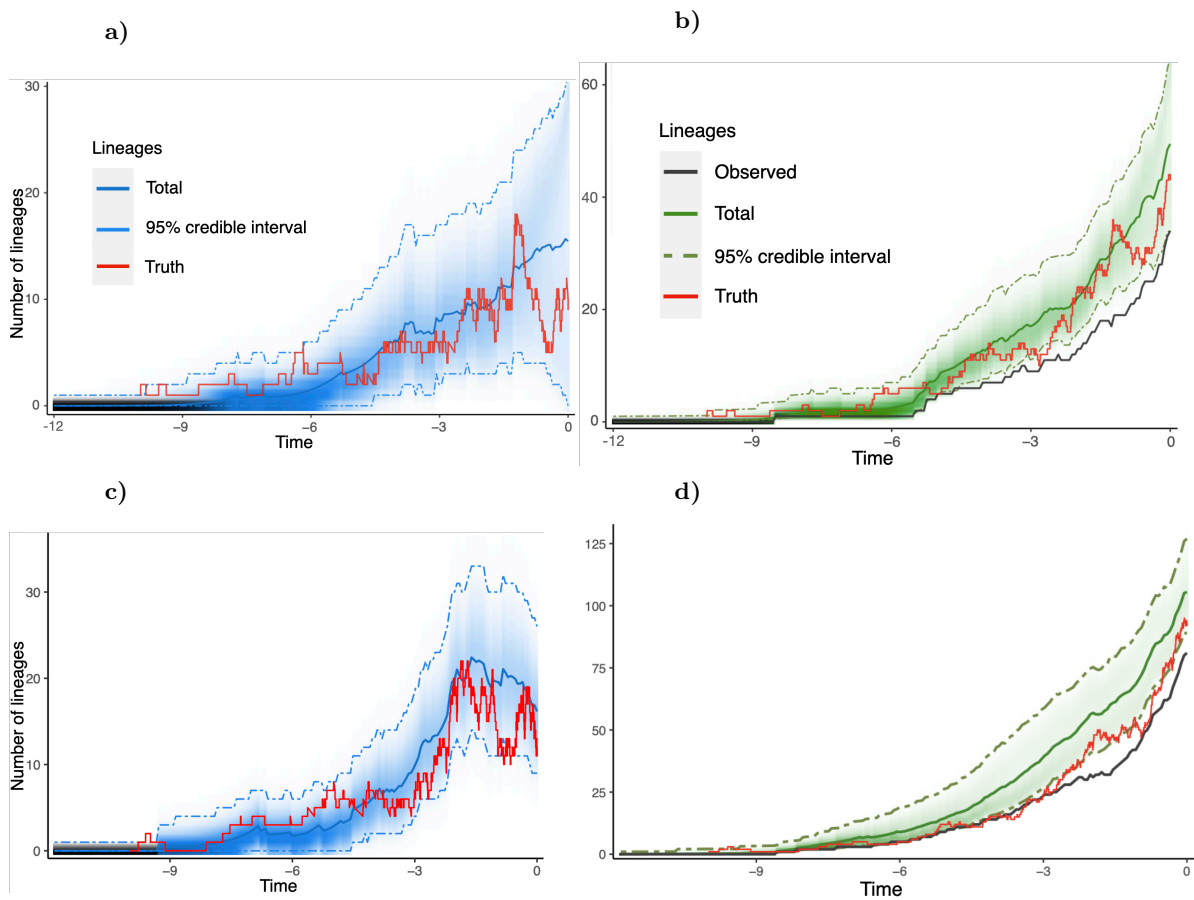
Figure S8. Validation of the diversity dynamics inferred by OBDP compared to the true simulated data. a) Posterior probability distribution of the number of hidden lineages through time for "dataset 1", plotted with the new RevGadgets utilities. b) Posterior probability distribution of the total number of lineages through time for "dataset 1". c-d) Same as a-b), but for "dataset 2". The 95% credible intervals are indicated in dashed lines, the expected number of lineages is in blue or green and the true, simulated, trajectory in red. The black line represents the inferred Lineages Through Time (LTT) plot, note that the total diversity equals the LTT plus the hidden diversity.
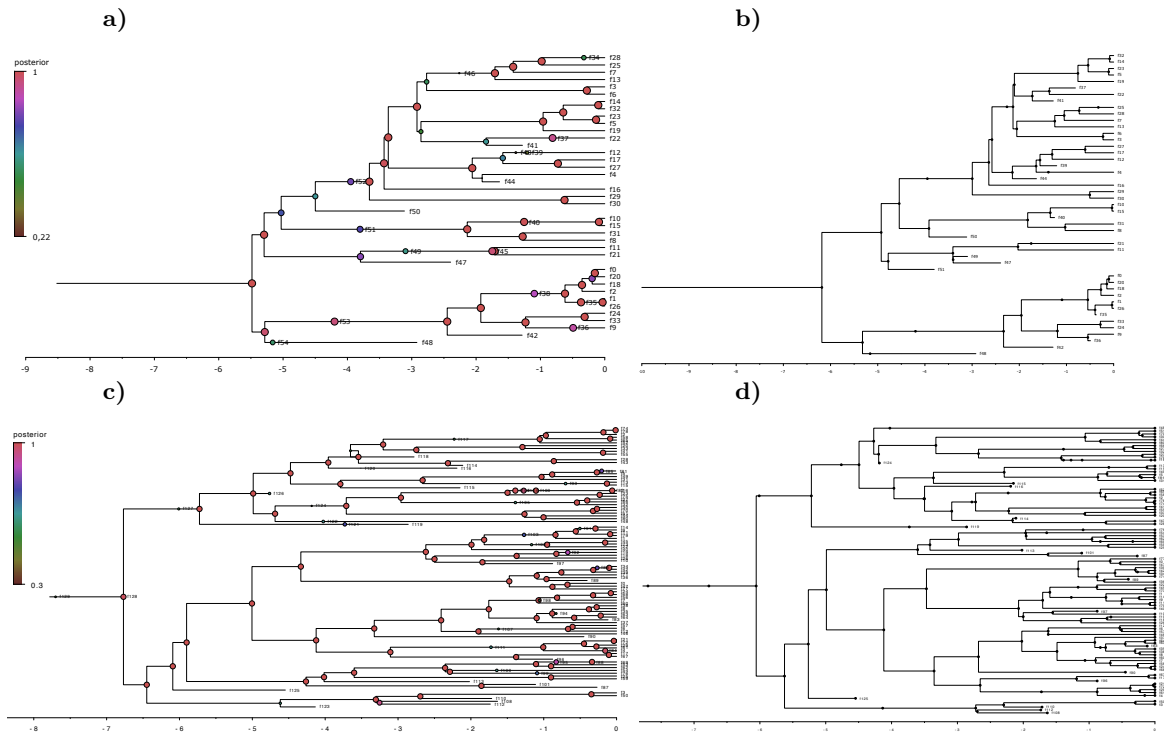
Figure S9. Validation of the inferred trees against the true simulated ones. a) Inferred phylogenetic tree for "dataset 1", visualized in FigTree 1.4.4. The node colors refer to their posterior probability. b) Original simulated tree for "dataset 1", aligned on the same temporal scale. Note that the topology is well recovered but divergence dates do not always perfectly match. c-d) Same as a-b) but on "dataset 2". Due to a greater amount of data in genetic sequences of both past and extant individuals, the divergence dates tend to be better inferred.

²⁵⁴ E – Macroevolution application: Inferring past cetacean diversity

²⁵⁵ *Preliminary analysis of the cetacean occurrence fossil record*

²⁵⁶   A detailed notebook is available at

²⁵⁷ https://github.com/Jeremy-Andreoletti/Cetacea_PBDB_Occurrences to follow our

²⁵⁸ exploration of the cetacean dataset. We identified several biases in their fossil record, in

²⁵⁹ particular much more variable occurrence densities – defined as the number of occurrences

²⁶⁰ by unit of time in the stratigraphic range of a clade – than expected from our model (see

²⁶¹ Figure S10).

²⁶²   Since OBDP assumes that only one individual of a species will be sampled at a

²⁶³ time, we subsampled the dataset to aggregate all occurrences of the same taxon found in

²⁶⁴ the same geological formation. This subsampling also reduced the observed discrepancy in

²⁶⁵ occurrence densities. The final subsampled dataset was composed of 968 occurrences.
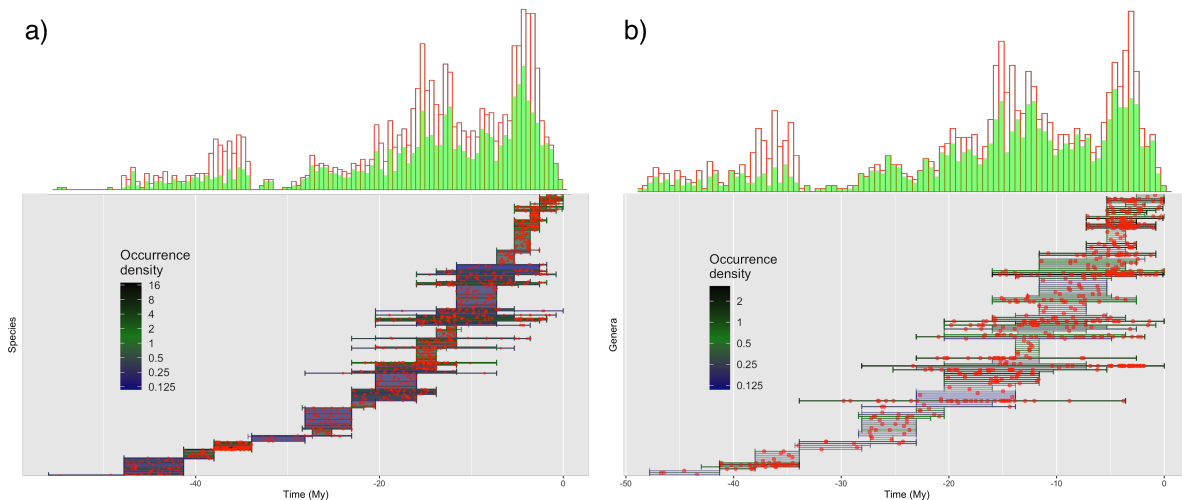
Figure S10. Occurrence distributions and bias correction, for cetacean species (a) and genera (b). At the top, occurrence distributions are compared before (red) and after (green) aggregating in geological formations. Below, stratigraphic ranges are displayed over time and colored according to the density of occurrences (red dots).

Table S6. Prior distributions for parameters and models of the Cetacea analysis. For each parameter its prior distribution, its initial value at the origin of the MCMC chain (set to speed up convergence) and the references that support these choices are indicated. Notations: $\mathcal{U}$ for the Uniform distribution, $\mathcal{E}$ for Exponential, $Log\mathcal{N}$ for Log-Normal, $\mathcal{G}$ for Gamma, Dir for Dirichlet, GTR for General Time Reversible and $JC69$ for the Jukes-Cantor 1969.

Component	Prior	Initial	Justification
t_{or}	$\mathcal{U}(max(occurrences), 60)$	$\frac{max+60}{2}$	Origin after the last occurrence. Initialised close to the estimated Whippomorpha root age from McGowen et al. (2020)
μ	$\mathcal{E}(5)$	0.05	Initialized according to estimations by Rabosky (2014)
$\lambda - \mu$	$Log\mathcal{N}(\ln[\frac{\ln 41}{t_{or}}], 0.587405)$	$\frac{\ln 41}{t_{or}}$	Expected number of species under a Birth-Death process centred around the observed number of genera. Lognormal distribution with 95% prior probability spanning exactly one order of magnitude (Höhna and Heath, 2019)
r	0	0	Removal probability at sampling, irrelevant in macroevolution
$\psi + \omega$	$\mathcal{E}(1)$	0.3	Unknown sampling rate for all fossils (including occurrences)
$\omega/(\psi + \omega)$	$\mathcal{U}(0, 1)$	Empirical	Unknown probability that morphological characters are available for a given fossil. Initialized at the empirical proportion of fossils with morphology among all fossils
Sampling bias	Messinian: $\mathcal{G}(2, 2)$ Aquitalian: $\mathcal{G}(2, 2)$ Rupelian: $\mathcal{G}(2, 2)$	0.75 0.5 0.1	Some geological stages are known to have transmitted a scarcer sedimentary record (Marx et al. 2016), thus fossil sampling rates are allowed to be estimated lower in these intervals.
ρ	$\mathcal{U}(0.95, 1)$	1	Sequences or morphology is used for the 41 accepted extant cetacean genera, but we allow for some still unknown genera
Fossil age uncertainty	$\mathcal{U}(min, max)$	Minimum age	Moves shifting a fossil age outside of its range are rejected (Heath et al., 2019)
Mean molecular clock rate	Nuclear: $\mathcal{U}(0, 0.01)$ Mitochondrial: $\mathcal{U}(0, 0.1)$	0.0.00075 0.03	Priors based on rates of molecular evolution for all mammals in Allio et al. (2017). Initialised at an intermediate rate between mysticetes and odontocetes as estimated by Dornburg et al. (2012).
Clock rate relaxation	Uncorrelated: $\mathcal{E}(1/mean)$	$mean$	Independent and identically distributed exponential rates are defined for each branch
Molecular substitution model: $GTR + \Gamma$	Exchangeability rates: $Dir(1, 1, 1, 1, 1, 1)$ Stationary frequencies: $Dir(1, 1, 1, 1)$ Gamma shape: $\mathcal{E}(1)$		Sophisticated nucleotide evolution model with rate variation across sites according to a discretized Gamma distribution. The Dirichlet distributions constrain vectors to sum to one (Heath et al., 2019)
Morphological substitution model: $JC69$	Strict clock rate: $\mathcal{E}(1)$ Gamma shape: $\mathcal{E}(1)$	0.5 0.125	Simpler character evolution model. Characters are partitioned according to their number of states (Wright, 2020)

Detailed priors used for Bayesian inference

We detail in Table S6 all priors used for the inference on the cetacean dataset.

Cetacean genera phylogeny

The Maximum Clade Credibility phylogeny was computed with RevBayes (Höhna et al., 2016), and plotted with Rstudio (RStudio Team, 2020) and the RevGadgets library (Tribble et al., 2021).
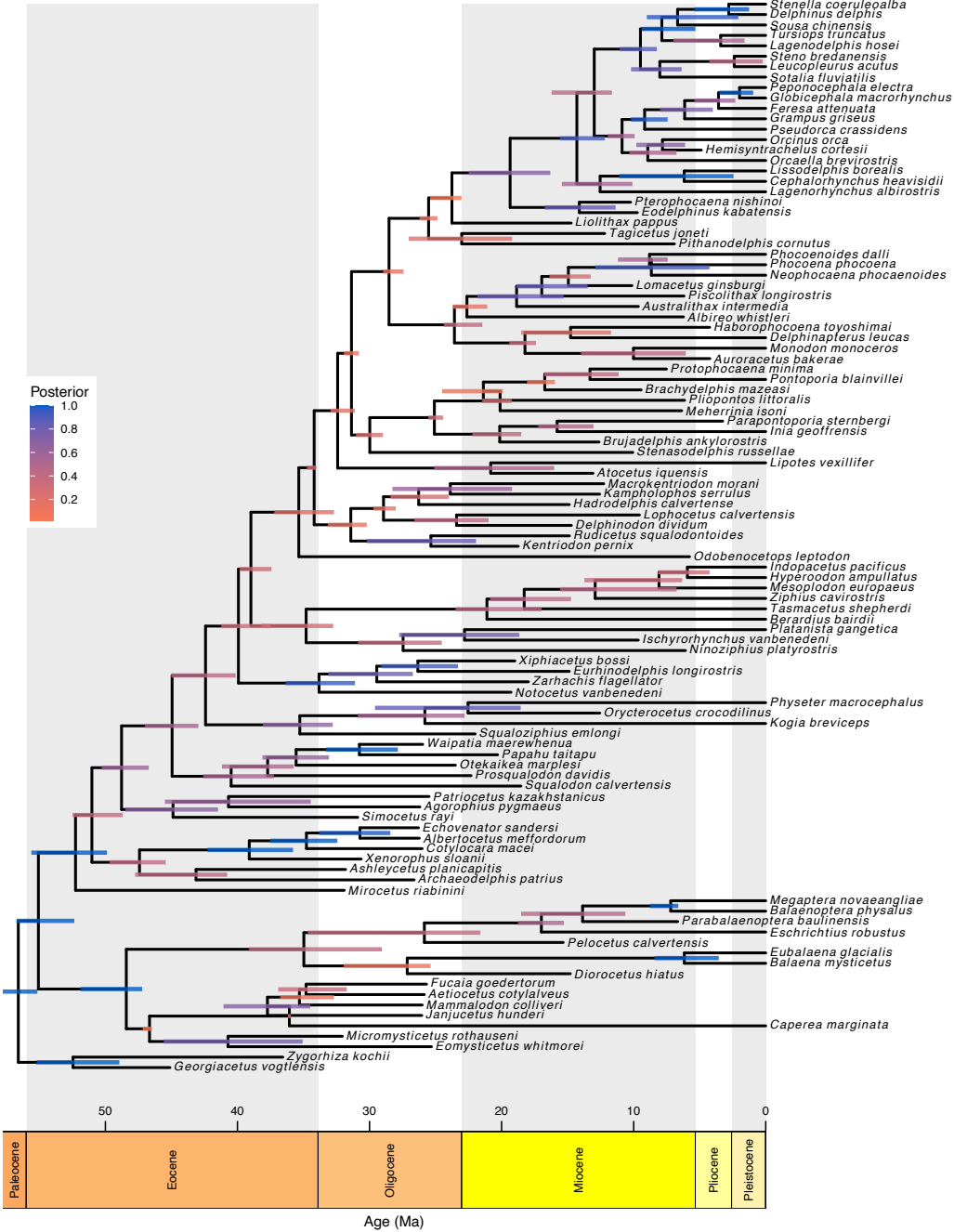
Figure S11. Maximum Clade Credibility phylogeny of the 41 currently accepted extant cetacean genera and 62 fossil genera. The colors of nodes bars reflect posterior probabilities.

F – Epidemiology application: the Diamond Princess SARS-2 COVID-19 outbreak dynamics

Data acquisition on GISAID

We gratefully acknowledge the following Authors from the Originating laboratories responsible for obtaining the specimens, as well as the Submitting laboratories where the genome data were generated and shared via GISAID, on which this research is based. All Submitters of data may be contacted directly via www.gisaid.org

accession ID EPI_ISL_416565, EPI_ISL_416566, EPI_ISL_416567, EPI_ISL_416568,

EPI_ISL_416569,

EPI_ISL_416570, EPI_ISL_416571, EPI_ISL_416572, EPI_ISL_416573,

EPI_ISL_416574, EPI_ISL_416575, EPI_ISL_416576, EPI_ISL_416577,

EPI_ISL_416578, EPI_ISL_416579, EPI_ISL_416580, EPI_ISL_416581,

EPI_ISL_416582, EPI_ISL_416583, EPI_ISL_416584, EPI_ISL_416585,

EPI_ISL_416586, EPI_ISL_416587, EPI_ISL_416588, EPI_ISL_416589,

EPI_ISL_416590, EPI_ISL_416591, EPI_ISL_416592, EPI_ISL_416593,

EPI_ISL_416594, EPI_ISL_416595, EPI_ISL_416596, EPI_ISL_416597,

EPI_ISL_416598, EPI_ISL_416599, EPI_ISL_416600, EPI_ISL_416601,

EPI_ISL_416602, EPI_ISL_416603, EPI_ISL_416604, EPI_ISL_416605,

EPI_ISL_416606, EPI_ISL_416607, EPI_ISL_416608, EPI_ISL_416609,

EPI_ISL_416610, EPI_ISL_416611, EPI_ISL_416612, EPI_ISL_416613,

EPI_ISL_416614, EPI_ISL_416615, EPI_ISL_416616, EPI_ISL_416617,

EPI_ISL_416618, EPI_ISL_416619, EPI_ISL_416620, EPI_ISL_416621,

EPI_ISL_416622, EPI_ISL_416623, EPI_ISL_416624, EPI_ISL_416625,

EPI_ISL_416626, EPI_ISL_416627, EPI_ISL_416628, EPI_ISL_416629,

EPI_ISL_416630, EPI_ISL_416631, EPI_ISL_416632, EPI_ISL_416633,

EPI_ISL_416634, EPI_ISL_454749

298 **Originating Laboratory** Japanese Quarantine Stations

299 **Submitting Laboratory** Pathogen Genomics Center, National Institute of Infectious

300 Diseases

301 **Authors** Tsuyoshi Sekizuka, Kentaro Itokawa, Rina Tanaka, Masanori Hashino, Tsutomu

302 Kageyama, Shinji Saito, Ikuyo Takayama, Hideki Hasegawa, Takuri Takahashi,

303 Hajime Kamiya, Takuya Yamagishi, Motoi Suzuki, Takaji Wakita, Makoto Kuroda

Figure S12. Genome sequences used, originating and submitting labs generated on GISAID. Content is reproduced above.

304 *Pre-processing the data*

305 All case count and sequencing data were available at a resolution of days.

306 In order to use the main method described in this article, the case count record had

307 to be pre-processed so that occurrences are spread throughout the days. For a day with a

308 case count of n newly infected individuals, we drew n time points uniformly distributed

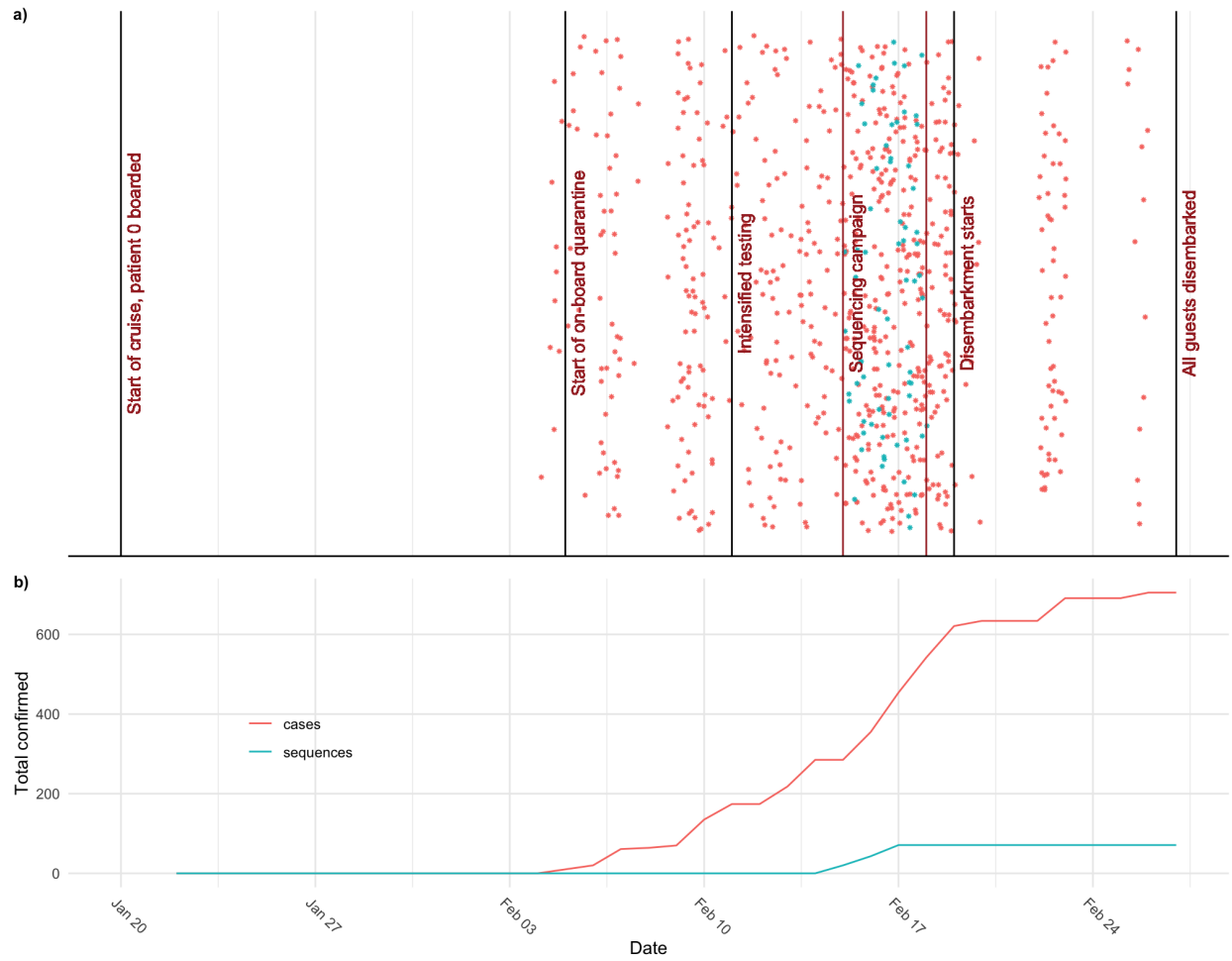309 throughout the day. The resulting dataset is shown in Figure S13.

Figure S13. Pre-processed dataset for the Diamond Princess outbreak analysis. a) Exact dates assigned to occurrences and sequences for the analysis. b) Total case counts and sequences through time.
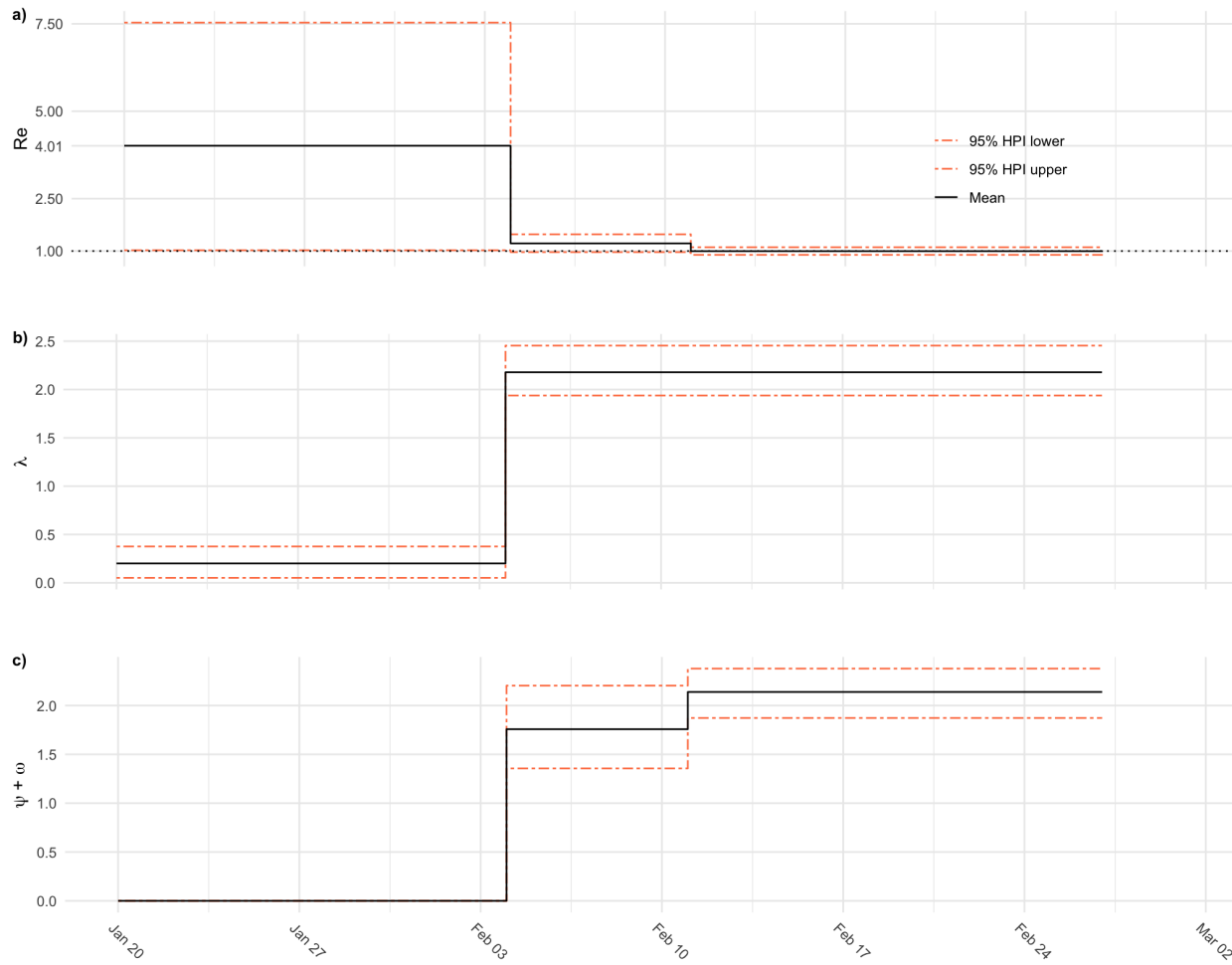
Figure S14. Detailed parameter estimates obtained from the COVID-19 outbreak analysis. a) Reproductive number estimates. b) Birth rate estimates. c) Total sampling (sequencing and PCR testing) rate estimates.

Detailed priors

We detail in Table S7 all priors used for the inference on the outbreak dataset of COVID-19 aboard the Diamond Princess.

The mean of the prior distribution of $\psi + \omega$ is set up to be the number of tests used on the ship, per day and per passenger, on the two periods.

- Within the first 7 days period, from February 4th to February 11th, there were 439 tests carried out, on 3711 passengers, leading to $\frac{439}{7 \times 3711} \approx 1.7 \times 10^{-2}$ tests per day per passenger.

- on the following 15 days period, from February 11th to February 27th, there were 3622 tests carried out, on 3711 passengers, leading to $\frac{3622}{15 \times 3711} \approx 6.5 \times 10^{-2}$ tests per day per passenger.

Table S7. Prior distributions for parameters and models of the SARS-2 COVID-19 analysis. For each parameter its prior distribution or value and the references that support these choices are indicated.

Component	Prior/Value	Shifts	Justification
t_{or}	38	N/A.	We study the outbreak from the start of the cruise on January 20, until February 27, when all guests were confirmed to have disembarked the ship, spanning a total period of 38 days. (https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/newpage_0032.html)
μ	$1/20$ day^{-1}	None.	In the absence of sampling and removal,infected individuals (patients) are assumed to become uninfectious on average 20 days after infection. (He et al., 2020)
λ	$\mathcal{U}(0, 24)$ $\mathcal{U}(0, 10)$	$t_m = (04.02.2020)$	The upper bound is set to 1 transmission per hour per infected individual before cabin isolation and lowered to 10 individuals after (maximal cabin size), from February 4th onward.
$\psi + \omega$	$Log\mathcal{N}\left(\frac{3622}{15*3711}, 0.5\right)$ $Log\mathcal{N}\left(\frac{439}{7*3711}, 0.5\right)$	$t_m = (11.02.2020, 04.02.2020)$	Testing started on February 4th and was intensified from February 11th onward, yielding two periods of 7 days and 15 days each. For each time period, the mean for the LogNormal distribution is set as the number of tests taken per passenger per day. The total numbers of tests carried out throughout the quarantine were communicated in press releases from the japanese Ministry of Health (https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/newpage_0032.html)
r	1	None.	Quarantine measures are assumed to have minimised contact between guests aboard. Patients testing positive were disembarked from the ship to a separate medical facility.
ρ	0	None.	No samples were sequenced after February 17th.
$\frac{\psi}{\omega + \psi}$	$\frac{71}{328}$	None.	Set to the fraction of the samples testing positive for COVID-19 that were sequenced.
Clock rate	8×10^{-4} substitutions per site per year	N/A.	Following Nexstrain (Hadfield et al., 2018).
Molecular substitution model: $GTR + \Gamma$	Exchangeability rates: $Dir(1, 1, 1, 1, 1, 1)$ Stationary frequencies: $Dir(1, 1, 1, 1)$ Gamma distribution shape: $\mathcal{E}(1)$		We allow for site rate heterogeneity, and assume unequal base frequencies and transition/transversion rates.