



An Alternative Approach to Compute the Likelihood for the DAISIE (Dynamic Assembly of Island Biota Through Speciation, Immigration, and Extinction) Model

Ornela N. Dehayem^{1,2}  · Bart Haegeman³  · Rampal S. Etienne¹ 

Received: 15 September 2025 / Accepted: 16 July 2026
© The Author(s) 2026

Abstract

Understanding island biodiversity requires studying the processes driving its assembly and examining how these processes vary over time, across lineages, and across different islands. The DAISIE (Dynamic Assembly of Island biota through Speciation, Immigration, and Extinction) framework has been developed for this purpose, and has advanced our understanding of island community assembly by estimating, from phylogenetic data, the contribution of the processes of colonization, speciation and extinction to island community assembly. However, the model assumes uniform colonization and diversification rates across lineages, ignoring potential variation in these rates caused by, for example, lineage-specific traits. This assumption thus restricts the model's capacity to capture complex colonization and diversification dynamics, and may consequently bias inference of colonization and diversification dynamics on islands. An extension of the framework behind these more complex dynamics is therefore desired. However, for this state-dependent colonization and diversification extension, the current computation of the likelihood of the model given phylogenetic data is computationally prohibitive. In this study, we therefore propose an alternative approach to computing the DAISIE likelihood, under the assumption of diversity-independent colonization and diversification. Our novel approach is based

✉ Ornela N. Dehayem
o.d.nanwou@rug.nl

✉ Rampal S. Etienne
r.s.etienne@rug.nl

Bart Haegeman
bart.haegeman@obs-banyuls.fr

¹ Groningen Institute for Evolutionary Life Sciences, University of Groningen, Box 11103, 9700 CC Groningen, The Netherlands

² Naturalis Biodiversity Center, Darwinweg 2, 2333 CR Leiden, The Netherlands

³ Laboratory of Microbial Oceanography, CNRS/Sorbonne Université, 1 avenue Pierre Fabre, 66650 Banyuls-sur-Mer, France

on the pruning algorithm that has been used for computing the likelihood of SSE (State-dependent Speciation and Extinction) models, which traces lineage history backward in time from the present day to the root of the tree. We demonstrate that our alternative approach reproduces DAISIE's predictions under diversity-independent colonization and diversification. This provides an additional layer of support to DAISIE in diversity-independent settings, but in doing so also gives more confidence in the results under diversity-dependence. Furthermore, our alternative approach offers greater computational efficiency. Finally, it provides a flexible framework for incorporating state-dependent dynamics in colonization, speciation, and extinction processes.

Keywords Macroevolution · Biodiversity patterns · Diversification model · Island biogeography · Phylogenetic trees

1 Introduction

Biodiversity, the variety of life forms on Earth, is a fundamental aspect of the natural world that underpins ecosystem function and resilience (Yadav et al. 2023; O'Connor et al. 2021). Understanding the processes that create and sustain biodiversity is a central goal in ecology and evolution. One of the primary approaches to studying the accumulation of biodiversity over evolutionary time scales is to reconstruct the past dynamics of the two evolutionary processes, speciation and extinction, that are thought to shape species richness over these time scales (Morlon et al. 2022; Morlon 2014). Diversification models fitted to time-calibrated phylogenies of extant species have been widely used to test evolutionary scenarios that may have shaped patterns of diversification through time (Nee et al. 1309; Maddison et al. 2007; Etienne et al. 2012; Etienne and Rosindell 2012; Morlon 2014; Condamine et al. 2013). Each model encodes a distinct scenario of speciation and extinction dynamics – either constant-rate, diversity-dependent, time-dependent, trait-dependent, or assuming protracted speciation. These hypotheses are evaluated by fitting a range of candidate models to identify the scenario that best approximates diversification patterns.

Marine islands present ideal systems to unravel evolutionary processes that have shaped present-day life due to their isolation, well-defined boundaries, and often exceptionally high biodiversity (Losos and Ricklefs 2009; Warren et al. 2015; Gillespie and Roderick 2014; Helmus et al. 2014; Patino et al. 2017). Their relatively recent geological origins and limited species richness compared to the much larger continental regions make studying processes shaping biodiversity more tractable. Moreover, islands (in an archipelago or across the globe) offer repeated natural experiments, providing multiple case studies for testing and contrasting evolutionary patterns across diverse environments, time and lineages (Losos and Ricklefs 2009). Island communities assemble through in situ speciation and extinction as well as via colonization (MacArthur and Wilson 2001). The DAISIE model (Dynamic Assembly of Island biota through Speciation, Immigration, and Extinction) has been developed to incorporate these three fundamental processes driving island diversity (Valente et al. 2015; Etienne et al. 2023). DAISIE makes use of information contained in phylogenies to

assess the role of the processes of colonization (arrival and establishment of new island species from a nearby continent), speciation via anagenesis (evolution to an island species distinct from the mainland ancestor due to reduced gene flow between island and mainland populations) and cladogenesis (the process of an ancestral species giving rise to two different species on the island or archipelago), and extinction (loss of species from the island) in shaping island communities. It also assesses whether the rates of colonization and cladogenesis depend negatively on island diversity, and hence limit diversity.

The DAISIE framework has been successfully applied to 1) quantify how colonization, speciation, and extinction rates vary with island age, area, and isolation (Valente et al. 2020); 2) assess the role of diversity limits and equilibrium processes in shaping patterns of species richness of birds in the Galápagos and Macaronesian islands (Valente et al. 2015, 2017); 3) estimate the amount of evolutionary time lost or threatened due to anthropogenic extinctions (Valente et al. 2017, 2019); 4) demonstrate how environmental changes, such as lake expansion, can elevate equilibrium diversity by enhancing colonization rates (Hauffe et al. 2020), and 5) study the phylogenetic limits to diversity-dependence (Etienne et al. 2023).

While the DAISIE framework has been instrumental in advancing our understanding of community assembly on islands, the current DAISIE model for which parameter estimation is possible through maximum likelihood has certain limitations. The likelihood computation currently requires uniform diversification and extinction parameters across lineages, thus ignoring variability associated with species-specific traits. However, species differ in dispersal abilities, physiological and morphological traits, habitat preferences and ecological interactions, all of which can modulate colonization frequency, speciation speed, and extinction risk. Ignoring such trait dependence can bias model inferences, especially if species' traits shape the structure of phylogenetic trees on islands (Xie et al. 2023; Maddison 2006; Maddison et al. 2007). Although DAISIE has been shown to effectively reconstruct biodiversity dynamics in many trait-dependent scenarios, it fails in some scenarios, notably when applied to phylogenetic data of community with large clade size variation (Xie et al. 2023). This limitation highlights the need for a trait-explicit modeling framework. However, extending the DAISIE framework to account for trait dependence is not straightforward, as the underlying mathematical formulation is not readily designed to incorporate trait-specific diversification and colonization dynamics. Here, we introduce an alternative methodology for modeling community assembly in isolated ecosystems under the assumption of diversity-independent diversification. Our approach follows the SSE (state-dependent speciation and extinction) framework introduced by Maddison et al. (2007), which operates backward in time, tracing lineage histories from the present day back to the island's origin. We demonstrate that our method aligns with DAISIE's predictions, while improving computational efficiency by making optimal use of the assumption on diversity-independence. Moreover, it provides a flexible framework that allows the incorporation of state-dependent colonization, speciation, and extinction processes. Our aim here is to present the framework rather than applying it in detail to a trait-dependent model, but we do provide an example of how our approach can be extended to incorporate trait-dependence.

2 Model and Likelihood Computation

2.1 Data Description and Assumptions

Our new methodology consists of an alternative approach to compute the likelihood of the DAISIE model as laid out in Valente et al. (2015). The DAISIE model assumes that after the emergence of an island, species can immigrate to the empty island from a mainland with a fixed species richness. Each colonization event initiates an independent lineage that may lead to 1) a single non-endemic species (same as the mainland ancestor), 2) a singleton endemic species (via anagenesis or cladogenesis followed by extinction), 3) a radiation of endemic species (via cladogenesis), or 4) extinction without leaving any extant descendants. The island community is thus represented by one or more such lineages, each corresponding to a separate colonization event and represented by its own phylogenetic tree. Figure 1 provides an example of a dataset for an island community. Each clade represents a separate colonization event from the mainland that has led to one or more species on the island. The dots represent the estimates of the colonization times, which are inferred by measuring the genetic divergence between island species and their mainland sister species (or mainland populations of the same species in the case of a non-endemic species). For the cases where these colonization times are unknown, the model integrates over all possible colonization times, that is, between the island-archipelago age and the present for singleton lineages (e.g., C and F), and between the island-archipelago age and the crown age for an endemic clade (e.g., A, B, D, E).

It is possible that before an observed colonization event, the same mainland species already colonized the island (Caujapé-Castells et al. 2017) (and persisted until the observed colonization event), but the model assumes that the observed lineage descends from the latest colonization event. This means that at the observed colonization event, any trace of these previous colonizations is erased (Valente et al. 2015). Recolonization of the same mainland ancestors can occur and still remain on the island, but only after the previous colonization has been followed by speciation. For this reason, the likelihood calculation assumes recolonization by the same mainland species is only possible after the island species undergoes speciation, because if this happened before speciation, the recolonizing population would have replaced the entire island population, and the most recent colonization event would be considered the new divergence time from the mainland which would be in disagreement with the data. Here we follow the same assumption; we do not question its validity in the current paper but will leave this for a future exploration. DAISIE can accurately estimate colonization and diversification rates within island communities in most cases from colonization times of independent lineages, their branching times, endemism status (derived from phylogenetic data), and the geological age of the island (Neves et al. 2021; Xie et al. 2023; Lambert et al. 2022). The model also allows for diversity-dependence in colonization and cladogenetic speciation to act either within clades (clade-specific, CS model) (Valente et al. 2015) or across the entire island community (island-wide, IW model) (Etienne et al. 2023). We note that for our methodology to work we assume that, in contrast to DAISIE, there is no diversity-dependent colonization or clado-

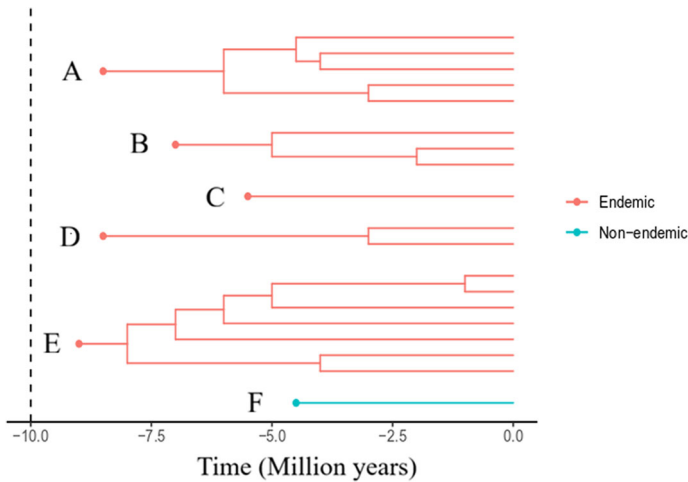


Fig. 1 Example of a dataset in the DAISIE framework. Each tree represents a colonization event with one or more descendants on the island. Circles indicate the time of colonization of each lineage; red - endemic lineages; blue - non-endemic lineages. The clades A, B, D, and E illustrate cases where a mainland species colonizes the island and subsequently gives rise to multiple descendant species. C corresponds to a scenario where a mainland species colonizes the island and speciates via anagenesis or cladogenesis, but only a single species survives until the present. F represents a scenario where a mainland species colonizes the island and does not undergo speciation. Time runs from the island age -10 (i.e., 10 million years ago) to the present time 0

genesis. Another key difference lies in how the two models handle incomplete data sampling. DAISIE uses n -sampling (Etienne et al. 2012), which assumes that, given a phylogenetic (or reconstructed) tree with a known number of extant species S , exactly n species are not sampled. In contrast, our approach uses the ρ -sampling scheme (FitzJohn et al. 2009), where the phylogeny is assumed to be a random sample of all extant species, where each species is included in the phylogeny with probability ρ , and with probability $1 - \rho$ to be unsampled. The two types of sampling are related by

$$P_{\rho}(\text{clade}) = \sum_{n=0}^{\infty} P_n(\text{clade}) \binom{S+n}{n} \rho^S (1-\rho)^n \quad (1)$$

where S is the number of species in the clade. In practice, ρ -sampling formulas are also applied to datasets consisting of a phylogeny with S tips together with a known number m of missing species. In that case, one typically sets $\rho = \frac{S}{S+m}$, corresponding to the observed fraction of sampled species. However, this should be seen as a pragmatic approximation rather than an exact equivalence to n -sampling.

2.2 Likelihood Computation

Mainland species are considered independent of one another, allowing the dynamics of each colonizing species to be modeled separately (CS model, see above). The

likelihood of observing the entire community is then calculated as the product of the probabilities of each independent colonization plus its potential descendants present on the island (which we will refer to as “lineage” hereafter even if it is a non-endemic singleton), and accounting for potential colonization events that occurred before and after the observed colonization, and subsequently went extinct. The model distinguishes four possible outcomes following a colonization event, the last three of which correspond to extant lineages observed on the island:

- No surviving lineage: a colonization event that resulted in a lineage that has gone completely extinct and is not represented among the present-day species on the island.
- Non-endemic singleton: a species present on the island that originated from a colonization event and has not undergone speciation since arrival, and is also found on the mainland.
- Endemic singleton: a species present on the island that originated from a colonization event and has undergone anagenetic or cladogenetic speciation, but only one species survives until the present on the island.
- Endemic radiation: a clade of two or more species that all descended from a single colonization event and underwent cladogenetic speciation on the island, resulting in a group of endemic species.

To compute the probabilities of observing a clade, DAISIE uses a hidden Markov formulation based on the variables $Q_n^k(t)$, which describe the probability that the diversification process is consistent with the observed phylogeny at time t , where k denotes the number of lineages represented in the phylogeny and n denotes the total number of additional (now extinct or unsampled) species present on the island at that time. The probabilities $Q_n^k(t)$ are computed forward in time, from the initial time t_0 , corresponding to the origin of the island, to the present time t_p .

This formulation becomes difficult to extend to trait-dependent models because it requires tracking not only the species represented in the phylogeny, but also the species that are not represented. To illustrate this, consider a time interval during which k species are represented in the phylogeny. If the trait has m possible states, the trait states of these k represented species can occur in m^k possible configurations. In addition, we must keep track of how many non-represented species occur in each of the m trait states. In principle the number of non-represented species is unbounded; however, we assume that it is bounded above by $n_{\max}$. Under this assumption, the number of possible configurations of non-represented species is $\binom{n_{\max}+m}{n_{\max}}$.

Thus, the total number of equations that must be tracked in a trait-dependent version of the current DAISIE framework scales as $m^k \binom{n_{\max}+m}{n_{\max}}$. This becomes computationally prohibitive as either the number of represented species or the number of trait states increases.

In contrast to the original DAISIE model, which traces lineage history forward in time from the island’s origin to the present, the approach we present in this paper traces it backward, from the present day to the island’s origin. Trait dependence can then be incorporated by indexing these quantities, and the associated rates, by trait state. This gives a much more compact system of equations, whose size is independent

of the number of represented species and scales linearly with the number of trait states.

We start by defining the backward time τ , measured from the present to the past. It is related to the usual forward time t by the equation $\tau = \tau_p - t + t_p$, where τ_p is the backward time at the present. We follow the differential equation-based approach introduced by Maddison et al. (2007), where probabilities are assigned at the tips of the lineage, and a set of rules is used to iteratively update these values as we trace the lineage backward in time, from the tips to the colonization time. At the colonization time of a lineage, we incorporate the probability of colonization, before continuing the computation further back in time to the island's origin.

We are now ready to define two central probabilities used in the model:

- $E(\tau)$, the probability that a lineage starting at time τ as an endemic species leaves no descendants at the present;
- $D(\tau)$, the probability that a lineage alive at time τ leaves a clade that is consistent with the observed data.

In order to capture the state of the system at specific points in time, we introduce several state-dependent variants of the probability $D(\tau)$. These states represent the probabilities of the dynamics being consistent with the observed data. We define these states as follows:

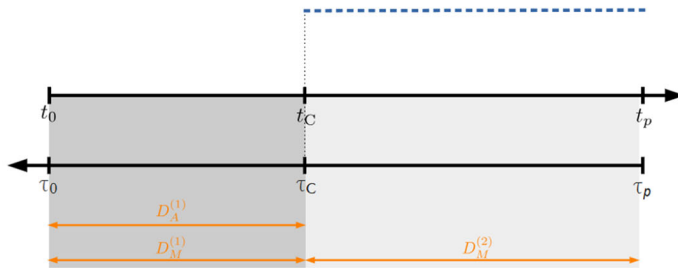
- $D_E(\tau)$ is the probability that a lineage that started on the island at time τ as an endemic lineage evolves into the observed clade;
- $D_M(\tau)$ is the probability that all lineages descending from the mainland species present at time τ , as well as those originating from subsequent colonizations, evolve in a manner consistent with the observed data;
- $D_A(\tau)$ is the probability that the mainland species is absent from the island at time τ , and that all lineages resulting from subsequent colonizations evolve in a manner consistent with the observed data.

Fig. 2 provides a schematic representation of the state probabilities $D_M(\tau)$ and $D_A(\tau)$ defined in the model.

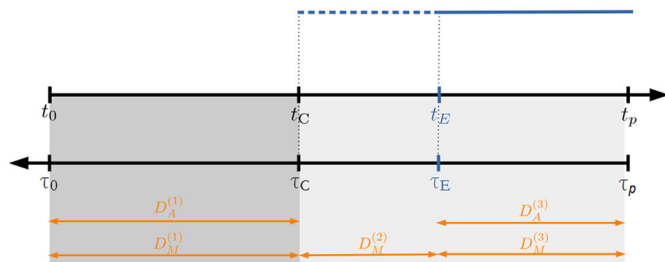
We construct the likelihoods by integrating a series of differential equations. To formulate these equations, it is convenient to introduce different instances of the variables D_A and D_M :

- $D_M^{(1)}(\tau)$, defined for $\tau \in [\tau_C, \tau_0]$, is the probability that the dynamics result in a clade consistent with the data, starting at time τ from an island with the non-endemic species which is not represented in the clade. This mainland species originates from a colonization event between τ_0 and τ_C . In this state the mainland species goes extinct before the present, or is taken over by the observed mainland species;
- $D_M^{(2)}(\tau)$, defined for $\tau \in [\tau_p, \tau_C]$, is the probability that the dynamics result in a clade consistent with the data, starting at time τ from an island with the non-endemic species which is represented in the clade. This mainland species originates from a colonization at τ_C . In this state the mainland species can speciate to initiate the observed lineage;
- $D_M^{(3)}(\tau)$, defined for $\tau \in [\tau_p, \tau_E]$, is the probability that the dynamics result in a clade consistent with the data, starting at time τ from an island with the non-endemic

(a) Non-endemic singleton



(b) Endemic singleton



(c) Endemic radiation

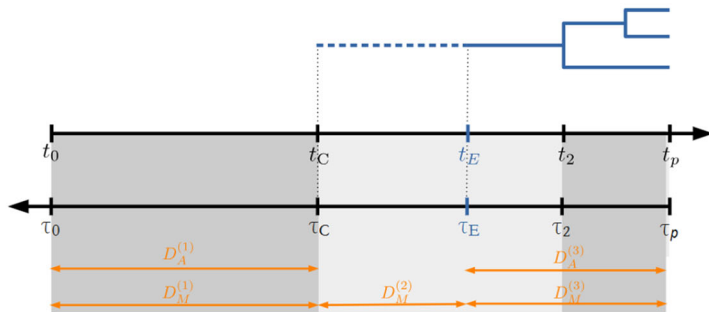


Fig. 2 Schematic representations of the temporal structure and definition of the state probabilities $D_M(\tau)$ and $D_A(\tau)$ in the backward pruning algorithm for different cases. The dashed blue line represents the time interval during which the species is non-endemic. The blue solid line represents the time interval during which the species is endemic. The orange arrows show the time intervals over which each probability is computed, going backward in time. The grey-shaded regions represent different time intervals, each governed by a distinct set of differential equations. These intervals correspond to different phases in the lineage history, reflecting changes in model dynamics across time. The times shown in black (τ_0 , τ_C , τ_2 , and τ_p) are extracted from the data. In contrast, the time points shown in blue (τ_E) are realization-dependent. These times are described as follows:

- t_0 , τ_0 : time of the island's origin, i.e. the island's age;
- t_C , τ_C : time of colonization event that gives rise to lineage observed at the present time;
- t_E , τ_E : time at which the endemic lineage at the present time becomes endemic;
- t_p , τ_p : present time

species which is not represented in the clade. This mainland species originates from a colonization after the speciation event on the observed lineage at τ_E . In this state the mainland species goes extinct before the present;

- $D_A^{(1)}(\tau)$, defined for $\tau \in [\tau_C, \tau_0]$, is the probability that the mainland species $D_M^{(1)}$ is absent from the island at time τ ;
- $D_A^{(3)}(\tau)$, defined for $\tau \in [\tau_p, \tau_E]$, is the probability that the mainland species $D_M^{(3)}$ is absent from the island at time τ .

Let us consider a lineage present on an island, with estimated colonization time τ_C , which may or may not have diversified into extant species. At the time τ_C the mainland ancestor colonized the island at a colonization rate γ . Following colonization, the lineage can diversify via cladogenesis at a rate λ^c , or anagenesis at a rate λ^a , or may go extinct at a rate μ . The likelihood of each clade is given by $D_A^{(1)}(\tau_0)$, which represent the probability that a process starting from an empty island at the time of island emergence τ_0 gives rise to the observed lineage on the island at time τ_p .

We now describe how to compute the likelihood for two specific cases: 1. no surviving species on the island and 2. a lineage with two or more species, assuming that the colonization time τ_C is known. The equations for singleton lineages can be derived analogously and are provided in Appendix A. The same framework can also be extended to the case where the colonization time is not precisely known, but is constrained to lie between $\tau_{\min}$ and $\tau_{\max}$. These extensions are also given in the Appendix A.

2.2.1 No Surviving Descendants

In this scenario, mainland species may have colonized, but none of the colonists survive to the present τ_p . The dynamics of the mainland species on the island are captured by the pair of variables $D_M^{(1)}$ and $D_A^{(1)}$. These state probabilities depend on the time τ (see Table 3). But for notational simplicity we omit this dependence in all equations throughout the manuscript. The dynamical equations describing the change in the probabilities from the present time τ_p to time of the island's origin τ_0 are given by:

$$\begin{aligned} \frac{dD_A^{(1)}}{d\tau} &= -\gamma D_A^{(1)} && \text{(no recolonization)} \\ &+ \gamma D_M^{(1)} && \text{(recolonization by mainland species)} \end{aligned} \quad (2a)$$

$$\begin{aligned} \frac{dD_M^{(1)}}{d\tau} &= -(\lambda^a + \lambda^c + \mu + \gamma) D_M^{(1)} && \text{(no anagenesis/cladogenesis/extinction/colonisation)} \\ &+ (\lambda^a E + \lambda^c E^2 + \mu) D_A^{(1)} && \text{(anagenesis followed by extinction; cladogenesis followed by extinction of both daughter lineages; extinction of mainland species; immediately after event mainland species is absent)} \\ &+ \gamma D_M^{(1)} && \text{(recolonisation by mainland species)} \end{aligned} \quad (2b)$$

$$\begin{aligned}
\frac{dE}{d\tau} &= \mu && \text{(extinction)} \\
&- (\lambda^c + \mu)E && \text{(no cladogenesis/extinction)} \\
&+ \lambda^c E^2 && \text{(cladogenesis followed by two extinctions)}
\end{aligned} \tag{2c}$$

In the equation for $D_M^{(1)}$, the term $+\gamma D_M^{(1)}$ cancels with the recolonisation component $-\gamma D_M^{(1)}$ contained in $-(\lambda^a + \lambda^c + \mu + \gamma)D_M^{(1)}$.

These equations describe the scenario in which lineages colonize the island at some point between τ_p and τ_0 , may or may not undergo speciation through anagenesis or cladogenesis, but all resulting lineages eventually go extinct before the present.

At the present time τ_p , the initial conditions are:

$$D_A^{(1)}(\tau_p) = 1, \quad D_M^{(1)}(\tau_p) = 0, \quad E(\tau_p) = 0. \tag{3}$$

This initialization reflects the fact that the mainland species is not observed on the island at the present time. Integrating these quantities backward to τ_0 gives the likelihood

$$\mathcal{L}_0 = D_A^{(1)}(\tau_0), \tag{4}$$

2.2.2 Endemic Non-singleton Lineage, with Colonization Time at τ_C

We consider an endemic non-singleton lineage whose mainland ancestor colonises the island at time τ_C . After colonisation, the lineage diversifies on the island and gives rise to several extant descendant species observed at the present time τ_p . The crown age of the endemic clade is denoted by τ_2 . After the first speciation of the lineage, the mainland ancestor may recolonize the island and go extinct again; this is captured by the variables $D_M^{(3)}$ and $D_A^{(3)}$.

The likelihood is computed by integrating backward in time from the present τ_p towards the island age τ_0 . We divide the history of the lineage into three intervals:

$$[\tau_p, \tau_2], \quad [\tau_2, \tau_C], \quad [\tau_C, \tau_0].$$

Dynamics between $[\tau_p, \tau_2]$

Between the crown age τ_2 and the present τ_p , the lineage is present, and has already undergone cladogenesis speciation. The dynamical equations describing the change in the probabilities between τ_p and τ_2 are given by:

$$\begin{aligned}
\frac{dD_E}{d\tau} &= -(\lambda^c + \mu)D_E && \text{(no cladogenesis/extinction)} \\
&+ 2\lambda^c D_E E && \text{(cladogenesis followed by extinction of one daughter lineage)}
\end{aligned} \tag{5a}$$

$$\begin{aligned} \frac{dD_M^{(3)}}{d\tau} = & -(\lambda^a + \lambda^c + \mu + \gamma)D_M^{(3)} && \text{(no anagenesis/cladogenesis/extinction/colonisation)} \\ & + (\lambda^a E + \lambda^c E^2 + \mu)D_A^{(3)} && \text{(anagenesis followed by extinction; cladogenesis followed by two extinctions; extinction of mainland species; immediately after event mainland species is absent)} \\ & + \gamma D_M^{(3)} && \text{(recolonisation by mainland species)} \end{aligned} \quad (5b)$$

$$\begin{aligned} \frac{dD_A^{(3)}}{d\tau} = & -\gamma D_A^{(3)} && \text{(no recolonisation)} \\ & + \gamma D_M^{(3)} && \text{(recolonisation by mainland species)} \end{aligned} \quad (5c)$$

$$\begin{aligned} \frac{dE}{d\tau} = & \mu && \text{(extinction)} \\ & - (\lambda^c + \mu)E && \text{(no cladogenesis/extinction)} \\ & + \lambda^c E^2 && \text{(cladogenesis followed by two extinctions)} \end{aligned} \quad (5d)$$

In the equation for $D_M^{(3)}$, the term $+\gamma D_M^{(3)}$ cancels with the recolonisation component $-\gamma D_M^{(3)}$ contained in $-(\lambda^a + \lambda^c + \mu + \gamma)D_M^{(3)}$. The initial conditions at τ_p are given by:

$$D_E(\tau_p) = 1, \quad D_M^{(3)}(\tau_p) = 0, \quad D_A^{(3)}(\tau_p) = 1, \quad E(\tau_p) = 0 \quad (6)$$

If species in the lineages are sampled at random (all species in the clade have the same probability of being sampled), incomplete sampling can be accounted for by modifying the initial conditions for each tip in the tree to reflect ρ , the probability of sampling a species, following FitzJohn et al. (2009). The initial conditions at τ_p are modified to:

$$D_E(\tau_p) = \rho, \quad D_M^{(3)}(\tau_p) = 0, \quad D_A^{(3)}(\tau_p) = 1, \quad E(\tau_p) = 1 - \rho \quad (7)$$

where ρ is the sampling fraction of the lineage.

The set of equations Eqs. (5a–5d) are integrated along the edges of the trees starting from the tips. At each branching time, the probabilities of the two daughter branches are combined by multiplying both probabilities with one another and with λ^c (from $\lambda^c d\tau$ which represents the probability of cladogenesis speciation over the small time interval $d\tau$). The resulting product serves as the initial condition for the ancestral (sub-tending) branch as we continue tracing the lineage backward toward its colonisation time.

Dynamics between $[\tau_2, \tau_C]$

In the time interval $[\tau_2, \tau_C]$, the species has already colonized and may or may not have speciated. We have to add the dynamical equation for $D_M^{(2)}$, because we know that

on at least a part of the branch the species is non-endemic. The dynamical equations describing the change in the probabilities between τ_2 and τ_C are given by:

$$\begin{aligned} \frac{dD_E}{d\tau} &= -(\lambda^c + \mu)D_E && \text{(no cladogenesis/extinction)} \\ &+ 2\lambda^c D_E E && \text{(cladogenesis followed by extinction of one} \\ &&& \text{daughter lineage; factor two corresponds to} \\ &&& \text{choice of surviving daughter lineage)} \end{aligned} \quad (8a)$$

$$\begin{aligned} \frac{dD_M^{(2)}}{d\tau} &= -(\lambda^a + \lambda^c + \mu + \gamma)D_M^{(2)} && \text{(no anagenesis/cladogenesis/extinction/colonization)} \\ &+ (\lambda^a D_E + 2\lambda^c D_E E)D_A^{(3)} && \text{(anagenesis; cladogenesis followed by extinction} \\ &&& \text{of one daughter lineage; immediately after event} \\ &&& \text{mainland species is absent)} \end{aligned} \quad (8b)$$

$$\begin{aligned} \frac{dD_M^{(3)}}{d\tau} &= -(\lambda^a + \lambda^c + \mu + \gamma)D_M^{(3)} && \text{(no anagenesis/cladogenesis/extinction/colonisation)} \\ &+ (\lambda^a E + \lambda^c E^2 + \mu)D_A^{(3)} && \text{(anagenesis followed by extinction; cladogenesis} \\ &&& \text{followed by two extinctions; extinction of} \\ &&& \text{mainland species; immediately after event} \\ &&& \text{mainland species is absent)} \\ &+ \gamma D_M^{(3)} && \text{(recolonisation by mainland species)} \end{aligned} \quad (8c)$$

$$\begin{aligned} \frac{dD_A^{(3)}}{d\tau} &= -\gamma D_A^{(3)} && \text{(no recolonization)} \\ &+ \gamma D_M^{(3)} && \text{(recolonization by mainland species)} \end{aligned} \quad (8d)$$

The equation is E are given by Eq. (2c)

As initial conditions for $D_E(\tau_2^+)$ at τ_2 , we take λ^c times the product of the D_E functions of the merging subclades. For $D_M^{(2)}(\tau_2^+)$ we use the same quantity multiplied by the probability $D_A^{(3)}(\tau_2^-)$ that the mainland species is not present on the island. This covers both cases in which the speciation event that transforms the mainland species into an endemic one occurs either between τ_C and τ_2 or precisely at time τ_2 . The initial conditions at τ_2 are thus:

$$D_E(\tau_2^+) = \lambda^c D_E^{(l)}(\tau_2^-)D_E^{(r)}(\tau_2^-) \quad (9a)$$

$$D_M^{(2)}(\tau_2^+) = \lambda^c D_E^{(l)}(\tau_2^-)D_E^{(r)}(\tau_2^-)D_A^{(3)}(\tau_2^-), \quad (9b)$$

$$D_M^{(3)}(\tau_2^+) = D_M^{(3)}(\tau_2^-), \quad (9c)$$

$$D_A^{(3)}(\tau_2^+) = D_A^{(3)}(\tau_2^-), \quad (9d)$$

$$E(\tau_2^+) = E(\tau_2^-) \quad (9e)$$

Here $D_E^{(l)}(\tau_2^-)$ and $D_E^{(r)}(\tau_2^-)$ are the probabilities that the two descendant lineages (left and right lineage in the phylogeny) are endemic just before τ_2 .

Dynamics between $[\tau_C, \tau_0]$

In the interval $[\tau_C, \tau_0]$ there is no clade, and so the mainland species is not represented. Still it could have colonized the island, a possibility we describe with the variable $D_M^{(1)}$. The dynamical equations describing the change in the probabilities between τ_C and τ_0 are given by Eqs. (2a–2c). The initial conditions at τ_C are given by:

$$D_A^{(1)}(\tau_C^+) = D_M^{(1)}(\tau_C^+) = \gamma D_M^{(2)}(\tau_C^-), \quad E(\tau_C^+) = E(\tau_C^-). \quad (10)$$

Integrating these quantities backward to τ_0 gives the likelihood:

$$\mathcal{L}_E = D_A^{(1)}(\tau_0), \quad (11)$$

Once the likelihood is written in terms of the backward variables D and E , extending the model from a non-trait-dependent to a trait-dependent setting requires indexing the rates and probabilities by the trait state. Thus, the constant rates λ^a , λ^c , μ , and γ are replaced by state-dependent rates, and additional transition terms q_{ij} are added to account for changes from trait state i to trait state j . To illustrate this extension, we present the trait-dependent analogue of the simplest case, i.e. the one without surviving descendants. Specifically, we consider a mainland species with trait state i_M has not colonized the island, or it has, but after establishment and potential speciation or trait change, the species or any of its descendants did not survive to the present time τ_p . The dynamical equations describing the change in the probabilities between the present time τ_p and the island age τ_0 are given by:

$$\begin{aligned} \frac{dD_{A,i_M}^{(1)}}{d\tau} &= -\gamma_M D_{A,i_M}^{(1)} && \text{(no recolonisation)} \\ &+ \gamma_M D_{M,i}^{(1)} && \text{(recolonisation of the mainland species)} \end{aligned} \quad (12a)$$

$$\begin{aligned} \frac{dD_{M,i}^{(1)}}{d\tau} &= -\left(\lambda_i^a + \lambda_i^c + \mu_i + \sum_{j \neq i} q_{ij}\right) D_{M,i}^{(1)} && \text{(no anagenesis, no cladogenesis, no} \\ &&& \text{extinction of the mainland species)} \\ &+ \left(\lambda_i^a E_i + \lambda_i^c E_i^2 + \mu_i\right. \\ &+ \sum_{j \neq i} q_{ij} p_{ij} E_j \left.) D_{A,i_M}^{(1)} \right. && \text{(anagenesis/cladogenesis/extinction/trait} \\ &&& \text{change followed by anagenesis)} \\ &+ \sum_{j \neq i} q_{ij} (1 - p_{ij}) D_{M,j}^{(1)} && \text{(trait change not followed by anagenesis)} \end{aligned} \quad (12b)$$

$$\frac{dE_i}{d\tau} = \mu_i \quad \text{(extinction of the lineage)}$$

$$\begin{aligned}
& - \left(\lambda_i^c + \mu_i + \sum_{j \neq i} q_{ij} \right) E_i && \text{(no cladogenesis, no extinction, no trait change)} \\
& + \lambda_i^c E_i^2 && \text{(cladogenesis followed by extinction of both daughter lineages)} \\
& + \sum_{j \neq i} q_{ij} E_j && \text{(trait change followed by extinction)} \quad (12c)
\end{aligned}$$

where iM represents the trait state of the mainland ancestor, the term $\sum_{j \neq i} q_{ij}$ the sum of all possible transition rates from state i to any other state j , with $j \neq i$ and p_{ij} represents the probability that a change from trait i to j causes anagenesis.

At the present time τ_p , the initial conditions are given by:

$$D_{A,iM}^{(1)}(\tau_p) = 1, \quad D_{M,i}^{(1)}(\tau_p) = 0, \quad E_i(\tau_p) = 0. \quad (13)$$

The likelihood is given by:

$$\mathcal{L}_0 = D_{A,iM}^{(1)}(\tau_0), \quad (14)$$

A complete and detailed derivation of the full trait-dependent extension of DAISIE-DE for all other scenarios will be presented in a separate paper.

We implemented these likelihood computations, including those for unknown colonization time in the R package DAISIE available on Github at <https://github.com/rsetienne/DAISIE/tree/master/R>, where we refer to them as DAISIE-DE, after the main equations for D and E .

3 Parameter Estimates by DAISIE vs DAISIE-DE

To evaluate the performance of our new approach DAISIE-DE and, as an important independent check of DAISIE, we compared its parameter estimates with those obtained from the DAISIE model across four empirical island biota datasets (Table 1) available in the package DAISIE. For each dataset, both models were fitted to estimate the rates of colonization, cladogenesis, anagenesis, and extinction. The carrying capacity was set to $K = \infty$ in DAISIE to eliminate the influence of diversity-dependent dynamics, allowing a comparison of the diversity-independent DAISIE and DAISIE-DE. Table 1 summarizes the resulting rate estimates. The comparison between the DAISIE and DAISIE-DE results shows that, for each dataset, the maximum likelihood and corresponding parameter estimates of the observed phylogenetic data are practically identical. The minute differences are due to numerical precision limits in integration and optimization.

It is important to note that all datasets used in this comparison were complete. When there are missing species, we cannot directly compare the likelihood we obtain from DAISIE-DE to that of DAISIE, because DAISIE uses a different kind of sampling (n -sampling where the number of unsampled species is specified, versus ρ -sampling which we use here, where the fraction of sampled species is specified). However, we show in Appendix A that we can compute the likelihood involving n -sampling from

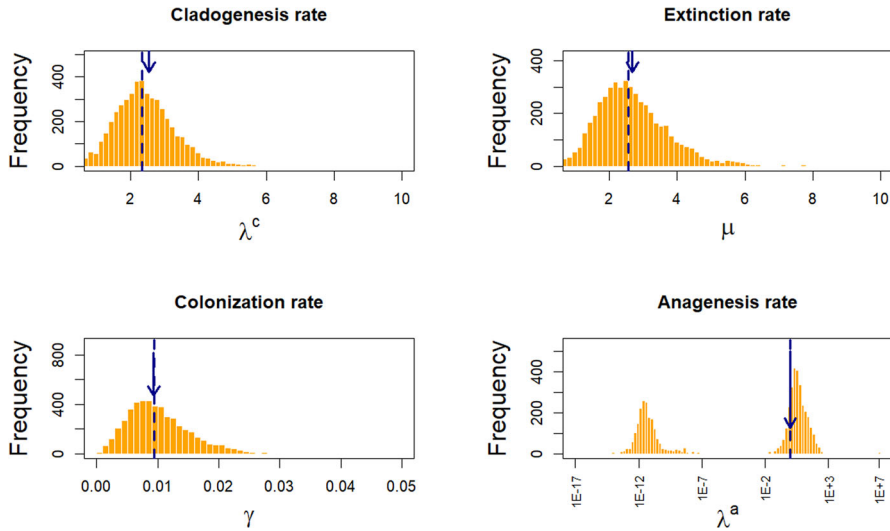


Fig. 3 Results of the parametric bootstrap analysis obtained by fitting the model to 5000 datasets simulated with the ML parameters of one of the 4 empirical datasets (the Galapagos phylogenetic dataset). Blue arrows indicate the true parameter values used in the simulations. The anagenesis rates are plotted on a logarithmic scale, and the distribution appears to be bimodal

our approach involving ρ -sampling, and we demonstrate with simulations that the results are practically identical to those from DAISIE and somewhat closer to the true values than with ρ -sampling (see Figures 5–7).

To further assess the robustness of DAISIE-DE's inference, we performed a parametric bootstrap analysis using one of the four empirical datasets, the Galapagos birds dataset. We fitted the model to the Galapagos birds dataset, and we simulated 5000 datasets under the model using the ML estimated parameters from the Galapagos birds dataset. We then re-estimated parameters for each simulated dataset and compared how much these estimates vary across the 5000 replicates. Figure 3 shows the distribution of parameter estimates across the 5000 replicates. This figure shows that the median values of the estimated rates of cladogenesis, colonization, anagenesis and extinction are close to the true generating values. However, the error and variance for anagenesis are noticeably larger than for other parameters. Estimating anagenesis was already challenging under the original DAISIE model (Valente et al. 2018, 2015). This is because anagenesis is primarily informed by endemic singleton lineages, but endemic singletons can also be produced by cladogenetic diversification followed by extinction.

4 Discussion

In this paper, we introduced an alternative approach for computing the likelihood of island-community assembly through colonisation, speciation, and extinction. The main motivation behind this reformulation was to facilitate the incorporation of

Table 1 Comparison of rates estimated by DAISIE and DAISIE-DE across different empirical datasets

Empirical dataset	S/C	Method	Time in s	λ^c	μ	γ	λ^a	logLik
Galápagos birds	25/8	DAISIE	45	2.5504	2.6835	0.00934	1.0073	-75.99
(Valente et al. 2015)		DAISIE-DE	18	2.5506	2.6838	0.00934	1.0072	-75.99
Hispaniola frogs	65/5	DAISIE	38	0.1787	0.0247	0.00062	0.000	-209.85
(Etienne et al. 2023)		DAISIE-DE	16	0.1787	0.0246	0.00062	0.000	-209.85
Cape Verde birds	10/10	DAISIE	196	0.0000	0.7678	0.0262	0.4628	-62.63
(Valente et al. 2019)		DAISIE-DE	16	0.0000	0.7676	0.0262	0.4628	-62.63
Biwa fish	71/68	DAISIE	1786	0.0801	1.4218	0.3917	0.3131	-230.54
(Hauffe et al. 2020)		DAISIE-DE	26	0.0801	1.4218	0.3917	0.3130	-230.54

The column "S/C" indicates the total number of species and colonisation events, respectively. The four empirical datasets are available in the package DAISIE

Table 2 Comparison of the approximate number of equations required by a trait-dependent extension of DAISIE, using the Q -approach, and the trait-dependent extension of the DAISIE-DE formulation, presented in this paper. Here, m is the number of trait states, k is the number of species represented in the phylogeny during the time interval considered, and $n_{\max} = 30$ is the maximal number of non-represented species that can be accounted for in the Q -approach, that is, it is the length of the Q -vector. For the Q -approach, the number of equations scales as $m^k \binom{n_{\max}+m}{n_{\max}}$. For the DAISIE-DE approach, we count the largest time segment as containing approximately $9m + 3$ trait-indexed equations

Number of trait states m	Represented species k	Q -approach equations	DAISIE-DE equations
2	5	1.6×10^4	21
2	10	5.1×10^5	21
2	20	5.2×10^8	21
4	5	4.7×10^7	39
4	10	4.9×10^{10}	39
4	20	5.1×10^{16}	39

additional biological complexity, such as trait dependence in colonisation and diversification rates. Such extensions are difficult to implement within the standard DAISIE likelihood framework, which is based on the Q -approach.

Table 2 compares the number of equations required under the trait-dependent extension of DAISIE, using the Q -approach, and under the trait-dependent DAISIE-DE formulation, using the backward approach.

To evaluate the performance of our new approach DAISIE-DE, we applied it to four empirical island biota datasets and compared its estimates of colonization, speciation (cladogenesis, anagenesis), and extinction rates with those obtained using the original DAISIE model (see Table 1). We found a strong agreement between the two models across all datasets, indicating that DAISIE-DE preserves the inferential robustness of DAISIE. In addition, DAISIE-DE is also computationally more efficient. The new likelihood calculation requires less computational time than the original DAISIE implementation. This is particularly important as the new diversity-independent model can be applied to much larger trees, making DAISIE-DE a practical alternative in these cases. Of course, one has to bear in mind that it cannot be used when colonization or diversification are diversity-dependent. More importantly, DAISIE-DE is built upon the same framework as State-dependent Speciation and Extinction (SSE) models, which opens the door to future extensions, such as incorporating trait-dependent diversification and colonization. In such a trait-dependent extension of the model, colonization, speciation, and extinction rates, as well as the probabilities associated with lineage states, would depend explicitly on species trait states, which themselves may change over time. We will explore this extension in future work. Furthermore, the similarity between DAISIE and DAISIE-DE lends support to the validity of the original DAISIE model, itself as two different approaches yield identical results across diverse empirical datasets.

Table 3 Parameters and probability/state variables used in the DAISIE-DE likelihood computation

Symbol / term	Meaning
<i>Model parameters (rates)</i>	
λ^c	Cladogenesis rate
μ	Extinction rate
λ^a	Anagenesis rate
γ	Colonisation rate
<i>State variables (probabilities)</i>	
$E(\tau)$	Probability that a lineage starting at time τ as an endemic species leaves no descendants at the present
$D_E(\tau)$	Probability that a lineage that started on the island at time τ as an endemic lineage evolves into the observed clade
$D_A^{(1)}(\tau)$	Probability that the mainland species corresponding to $D_M^{(1)}$ is absent from the island at time τ
$D_A^{(3)}(\tau)$	Probability that the mainland species corresponding to $D_M^{(3)}$ is absent from the island at time τ
$D_M^{(1)}(\tau)$	Probability that the dynamics evolve into a clade consistent with the data, starting at time τ from an island containing a non-endemic species that is not represented in the clade; this mainland species originates from a colonisation event between t_0 and t_C , and in this state it goes extinct before the present or is replaced by the observed mainland species
$D_M^{(2)}(\tau)$	Probability that the dynamics result in a clade consistent with the data, starting at time τ from an island containing a non-endemic species that is represented in the clade; this mainland species originates from a colonisation at τ_C and can speciate to initiate the observed lineage
$D_M^{(3)}(\tau)$	Probability that the dynamics lead to a clade consistent with the data, starting at time τ from an island containing a non-endemic species that is not represented in the clade; this mainland species originates from a colonisation after the speciation event at τ_E and goes extinct before the present

Appendix A Likelihood of Singleton Lineages and Lineages with Unknown Colonization Times

Here we present the likelihood calculation for singleton lineages and lineages whose estimated colonization time is unknown, and for which only an estimated maximum and/or minimum colonization time is available. Parameters and variables used in the DAISIE-DE likelihood computations are listed in Table 3, while the dynamics of all cases are summarized in Tables 4 and 5. Figure 4 provides a schematic representation of the state probabilities $D_M(\tau)$ and $D_A(\tau)$ defined in the model.

The times shown in black (τ_0 , τ_{max} , τ_{min} , τ_2 and τ_p) are extracted from the data. In contrast, the time points shown in blue ($\tau_{C,1}$, τ_C and τ_E) are realization-dependent. These times are described as follows:

Table 4 Differential equations defined for different cases with given colonization time τ_C over specific time intervals. The integration runs from right to left and the initial conditions in a column connect the equations to the equations to the left. The shading colors mean that the same set of equations is being used

Type of lineages	$[\tau_p, \tau_1]$	$[\tau_1, \tau_0]$
No surviving descendants	Initial conditions: $D_A^{(1)}(\tau_p) = 1, D_M^{(1)}(\tau_p) = 0, E(\tau_p) = 0$ System of equations: $\frac{dD_A^{(1)}}{d\tau} = -\gamma D_A^{(1)} + \gamma D_M^{(1)}$ $\frac{dD_M^{(1)}}{d\tau} = -(\lambda^a + \lambda^c + \mu)D_M^{(1)} + (\lambda^a E + \lambda^c E^2 + \mu)D_A^{(1)}$ $\frac{dE}{d\tau} = \mu - (\lambda^c + \mu)E + \lambda^c E^2$	
Non-endemic lineage (colonized at τ_1)	Initial conditions: $D_A^{(2)}(\tau_p) = 1, E(\tau_p) = 0$ System of equations: $\frac{dD_M^{(2)}}{d\tau} = -(\lambda^a + \lambda^c + \mu + \gamma)D_M^{(2)}$ $\frac{dE}{d\tau} = \mu - (\lambda^c + \mu)E + \lambda^c E^2$	Initial conditions: $D_A^{(1)}(\tau_1^+) = D_M^{(1)}(\tau_1^+) = \gamma D_M^{(2)}(\tau_1^-), E(\tau_1^+) = E(\tau_1^-)$ System of equations: $\frac{dD_A^{(1)}}{d\tau} = -\gamma D_A^{(1)} + \gamma D_M^{(1)}$ $\frac{dD_M^{(1)}}{d\tau} = -(\lambda^a + \lambda^c + \mu)D_M^{(1)} + (\lambda^a E + \lambda^c E^2 + \mu)D_A^{(1)}$ $\frac{dE}{d\tau} = \mu - (\lambda^c + \mu)E + \lambda^c E^2$
Endemic singleton lineage (colonized at t_1)	Initial conditions: $D_E(\tau_p) = 1, D_A^{(3)}(\tau_p) = 1, D_M^{(3)}(\tau_p) = 0,$ $D_M^{(2)}(\tau_p) = 0, E(\tau_p) = 0$ System of equations: $\frac{dD_E}{d\tau} = -(\lambda^c + \mu)D_E + 2\lambda^c D_E E$ $\frac{dD_M^{(2)}}{d\tau} = -(\lambda^a + \lambda^c + \mu + \gamma)D_M^{(2)} + (\lambda^a D_E + 2\lambda^c D_E E)D_A^{(3)}$ $\frac{dD_M^{(3)}}{d\tau} = -(\lambda^a + \lambda^c + \mu)D_M^{(3)} + (\lambda^a E + \lambda^c E^2 + \mu)D_A^{(3)}$ $\frac{dD_A^{(3)}}{d\tau} = -\gamma D_A^{(3)} + \gamma D_M^{(3)}$ $\frac{dE}{d\tau} = \mu - (\lambda^c + \mu)E + \lambda^c E^2$	Initial conditions: $D_A^{(1)}(\tau_1^+) = D_M^{(1)}(\tau_1^+) = \gamma D_M^{(2)}(\tau_1^-), E(\tau_1^+) = E(\tau_1^-)$ System of equations: $\frac{dD_A^{(1)}}{d\tau} = -\gamma D_A^{(1)} + \gamma D_M^{(1)}$ $\frac{dD_M^{(1)}}{d\tau} = -(\lambda^a + \lambda^c + \mu)D_M^{(1)} + (\lambda^a E + \lambda^c E^2 + \mu)D_A^{(1)}$ $\frac{dE}{d\tau} = \mu - (\lambda^c + \mu)E + \lambda^c E^2$
Endemic non-singleton lineage (colonized at τ_1)	Initial conditions: $D_E(\tau_p) = 1, D_A^{(3)}(\tau_p) = 1, D_M^{(3)}(\tau_p) = 0,$ $D_M^{(2)}(\tau_p) = 0, E(\tau_p) = 0$ System of equations: $\frac{dD_E}{d\tau} = -(\lambda^c + \mu)D_E + 2\lambda^c D_E E$ $\frac{dD_M^{(3)}}{d\tau} = -(\lambda^a + \lambda^c + \mu)D_M^{(3)} + (\lambda^a E + \lambda^c E^2 + \mu)D_A^{(3)}$ $\frac{dD_A^{(3)}}{d\tau} = -\gamma D_A^{(3)} + \gamma D_M^{(3)}$ $\frac{dE}{d\tau} = \mu - (\lambda^c + \mu)E + \lambda^c E^2$	Initial conditions: $D_A^{(1)}(\tau_1^+) = D_M^{(1)}(\tau_1^+) = \gamma D_M^{(2)}(\tau_1^-), E(\tau_1^+) = E(\tau_1^-)$ System of equations: $\frac{dD_A^{(1)}}{d\tau} = -\gamma D_A^{(1)} + \gamma D_M^{(1)}$ $\frac{dD_M^{(1)}}{d\tau} = -(\lambda^a + \lambda^c + \mu)D_M^{(1)} + (\lambda^a E + \lambda^c E^2 + \mu)D_A^{(1)}$ $\frac{dE}{d\tau} = \mu - (\lambda^c + \mu)E + \lambda^c E^2$

Table 5 Differential equations defined for different cases with maximum and minimum colonization times over specific time intervals. The integration runs from right to left and the initial conditions in a column connect the equations to the equations to the left. The shading colors mean that the same set of equations is being used

Type of lineages	$[\tau_p, \tau_{\min}]$	$[\tau_{\min}, \tau_2]$	$[\tau_2, \tau_{\max}]$	$[\tau_{\max}, \tau_0]$
Non-endemic singleton with maximum age of colonization at $\tau_{\max}$	Initial conditions: $D_A^{(1)}(\tau_p) = 0, D_M^{(1)}(\tau_p) = 0, D_M^{(2)}(\tau_p) = 1, E(\tau_p) = 0.$ System of equations: $\frac{dD_A^{(2)}}{d\tau} = -\gamma D_A^{(2)} + \gamma D_M^{(2)}$ $\frac{dD_M^{(1)}}{d\tau} = -(\lambda^c + \lambda^c + \mu + \gamma) D_M^{(1)} + (\lambda^c E + \lambda^c E^2 + \mu) D_A^{(2)} + \gamma D_M^{(2)}$ $\frac{dD_M^{(2)}}{d\tau} = -(\lambda^c + \lambda^c + \mu) D_M^{(2)} + (\lambda^c E + \lambda^c E^2 + \mu) D_A^{(2)}$ $\frac{dE}{d\tau} = \mu - (\lambda^c + \mu) E + \lambda^c E^2$			Initial conditions: $D_A^{(1)}(\tau_{\max}) = D_A^{(2)}(\tau_{\max}), D_M^{(1)}(\tau_{\max}) = D_M^{(2)}(\tau_{\max}), E(\tau_{\max}) = E(\tau_{\max}).$ System of equations: $\frac{dD_A^{(1)}}{d\tau} = -\gamma D_A^{(1)} + \gamma D_M^{(1)}$ $\frac{dD_M^{(1)}}{d\tau} = -(\lambda^c + \lambda^c + \mu) D_M^{(1)} + (\lambda^c E + \lambda^c E^2 + \mu) D_A^{(1)}$ $\frac{dE}{d\tau} = \mu - (\lambda^c + \mu) E + \lambda^c E^2$
Endemic singleton lineage with maximum age of colonization at $\tau_{\max}$	Initial conditions: $D_A^{(1)}(\tau_p) = 1, D_A^{(2)}(\tau_p) = 0, D_A^{(3)}(\tau_p) = 1, D_M^{(1)}(\tau_p) = 0, D_M^{(2)}(\tau_p) = 0, D_M^{(3)}(\tau_p) = 0, E(\tau_p) = 0.$ System of equations: $\frac{dD_A^{(1)}}{d\tau} = -(\lambda^c + \mu) D_A^{(1)} + 2\lambda^c D_A^{(1)} E$ $\frac{dD_A^{(2)}}{d\tau} = -\gamma D_A^{(2)} + \gamma D_M^{(2)}$ $\frac{dD_A^{(3)}}{d\tau} = -\gamma D_A^{(3)} + \gamma D_M^{(3)}$ $\frac{dD_M^{(1)}}{d\tau} = -(\lambda^c + \lambda^c + \mu + \gamma) D_M^{(1)} + (\lambda^c E + \lambda^c E^2 + \mu) D_A^{(2)} + \gamma D_M^{(2)}$ $\frac{dD_M^{(2)}}{d\tau} = -(\lambda^c + \lambda^c + \mu) D_M^{(2)} + (\lambda^c E + \lambda^c E^2 + \mu) D_A^{(2)} + (\lambda^c D_A^{(1)} + 2\lambda^c D_A^{(1)} E) D_A^{(3)}$ $\frac{dD_M^{(3)}}{d\tau} = -(\lambda^c + \lambda^c + \mu) D_M^{(3)} + (\lambda^c E + \lambda^c E^2 + \mu) D_A^{(3)}$ $\frac{dE}{d\tau} = \mu - (\lambda^c + \mu) E + \lambda^c E^2$			Initial conditions: $D_A^{(1)}(\tau_{\max}) = D_A^{(2)}(\tau_{\max}), D_M^{(1)}(\tau_{\max}) = D_M^{(2)}(\tau_{\max}), E(\tau_{\max}) = E(\tau_{\max}).$ System of equations: $\frac{dD_A^{(1)}}{d\tau} = -\gamma D_A^{(1)} + \gamma D_M^{(1)}$ $\frac{dD_M^{(1)}}{d\tau} = -(\lambda^c + \lambda^c + \mu) D_M^{(1)} + (\lambda^c E + \lambda^c E^2 + \mu) D_A^{(1)}$ $\frac{dE}{d\tau} = \mu - (\lambda^c + \mu) E + \lambda^c E^2$
Endemic non-singleton lineage with maximum age of colonization at $\tau_{\max}$	Initial conditions: $D_A^{(1)}(\tau_p) = 1, E(\tau_p) = 0, D_A^{(2)}(\tau_p) = 1, D_M^{(3)}(\tau_p) = 0.$ System of equations: $\frac{dD_A^{(1)}}{d\tau} = -(\lambda^c + \mu) D_A^{(1)} + 2\lambda^c D_A^{(1)} E$ $\frac{dD_A^{(2)}}{d\tau} = -(\lambda^c + \lambda^c + \mu) D_A^{(2)} + (\lambda^c E + \lambda^c E^2 + \mu) D_A^{(3)}$ $\frac{dD_A^{(3)}}{d\tau} = -\gamma D_A^{(3)} + \gamma D_M^{(3)}$ $\frac{dE}{d\tau} = \mu - (\lambda^c + \mu) E + \lambda^c E^2$	Initial conditions: $D_A^{(1)}(\tau_p) = 1, D_A^{(2)}(\tau_p) = 0, D_A^{(3)}(\tau_p) = 1, D_M^{(3)}(\tau_p) = 0.$ System of equations: $\frac{dD_A^{(1)}}{d\tau} = -(\lambda^c + \mu) D_A^{(1)} + 2\lambda^c D_A^{(1)} E$ $\frac{dD_A^{(2)}}{d\tau} = -(\lambda^c + \lambda^c + \mu) D_A^{(2)} + (\lambda^c E + \lambda^c E^2 + \mu) D_A^{(3)}$ $\frac{dD_A^{(3)}}{d\tau} = -\gamma D_A^{(3)} + \gamma D_M^{(3)}$ $\frac{dE}{d\tau} = \mu - (\lambda^c + \mu) E + \lambda^c E^2$	Initial conditions: $D_A^{(1)}(\tau_p) = 1, D_A^{(2)}(\tau_p) = 0, D_A^{(3)}(\tau_p) = 1, D_M^{(3)}(\tau_p) = 0.$ System of equations: $\frac{dD_A^{(1)}}{d\tau} = -(\lambda^c + \mu) D_A^{(1)} + 2\lambda^c D_A^{(1)} E$ $\frac{dD_A^{(2)}}{d\tau} = -(\lambda^c + \lambda^c + \mu) D_A^{(2)} + (\lambda^c E + \lambda^c E^2 + \mu) D_A^{(3)}$ $\frac{dD_A^{(3)}}{d\tau} = -\gamma D_A^{(3)} + \gamma D_M^{(3)}$ $\frac{dE}{d\tau} = \mu - (\lambda^c + \mu) E + \lambda^c E^2$	Initial conditions: $D_A^{(1)}(\tau_{\max}) = D_A^{(2)}(\tau_{\max}), D_M^{(1)}(\tau_{\max}) = D_M^{(2)}(\tau_{\max}), E(\tau_{\max}) = E(\tau_{\max}).$ System of equations: $\frac{dD_A^{(1)}}{d\tau} = -\gamma D_A^{(1)} + \gamma D_M^{(1)}$ $\frac{dD_M^{(1)}}{d\tau} = -(\lambda^c + \lambda^c + \mu) D_M^{(1)} + (\lambda^c E + \lambda^c E^2 + \mu) D_A^{(1)}$ $\frac{dE}{d\tau} = \mu - (\lambda^c + \mu) E + \lambda^c E^2$
Non-endemic singleton lineage with maximum and minimum age of colonization	Initial conditions: $D_A^{(1)}(\tau_p) = 1, E(\tau_p) = 0$ System of equations: $\frac{dD_A^{(2)}}{d\tau} = -(\lambda^c + \lambda^c + \mu + \gamma) D_A^{(2)} + \gamma D_M^{(2)}$ $\frac{dE}{d\tau} = \mu - (\lambda^c + \mu) E + \lambda^c E^2$	Initial conditions: $D_A^{(1)}(\tau_{\min}) = 0, D_A^{(2)}(\tau_{\min}) = 0, D_A^{(3)}(\tau_{\min}) = 1, E(\tau_{\min}) = E(\tau_{\min}).$ System of equations: $\frac{dD_A^{(2)}}{d\tau} = -\gamma D_A^{(2)} + \gamma D_M^{(2)}$ $\frac{dD_A^{(3)}}{d\tau} = -(\lambda^c + \lambda^c + \mu + \gamma) D_A^{(3)} + (\lambda^c E + \lambda^c E^2 + \mu) D_A^{(2)} + \gamma D_M^{(2)}$ $\frac{dD_M^{(2)}}{d\tau} = -(\lambda^c + \lambda^c + \mu) D_M^{(2)} + (\lambda^c E + \lambda^c E^2 + \mu) D_A^{(2)}$ $\frac{dE}{d\tau} = \mu - (\lambda^c + \mu) E + \lambda^c E^2$	Initial conditions: $D_A^{(1)}(\tau_{\max}) = 0, D_A^{(2)}(\tau_{\max}) = 0, D_A^{(3)}(\tau_{\max}) = 1, E(\tau_{\max}) = E(\tau_{\max}).$ System of equations: $\frac{dD_A^{(1)}}{d\tau} = -\gamma D_A^{(1)} + \gamma D_M^{(1)}$ $\frac{dD_M^{(1)}}{d\tau} = -(\lambda^c + \lambda^c + \mu) D_M^{(1)} + (\lambda^c E + \lambda^c E^2 + \mu) D_A^{(1)}$ $\frac{dE}{d\tau} = \mu - (\lambda^c + \mu) E + \lambda^c E^2$	Initial conditions: $D_A^{(1)}(\tau_{\max}) = D_A^{(2)}(\tau_{\max}), D_M^{(1)}(\tau_{\max}) = D_M^{(2)}(\tau_{\max}), E(\tau_{\max}) = E(\tau_{\max}).$ System of equations: $\frac{dD_A^{(1)}}{d\tau} = -\gamma D_A^{(1)} + \gamma D_M^{(1)}$ $\frac{dD_M^{(1)}}{d\tau} = -(\lambda^c + \lambda^c + \mu) D_M^{(1)} + (\lambda^c E + \lambda^c E^2 + \mu) D_A^{(1)}$ $\frac{dE}{d\tau} = \mu - (\lambda^c + \mu) E + \lambda^c E^2$
Endemic singleton lineage with maximum and minimum age of colonization	Initial conditions: $D_A^{(1)}(\tau_p) = 1, D_A^{(2)}(\tau_p) = 0, D_A^{(3)}(\tau_p) = 1, E(\tau_p) = 0.$ System of equations: $\frac{dD_A^{(1)}}{d\tau} = -(\lambda^c + \mu) D_A^{(1)} + 2\lambda^c D_A^{(1)} E$ $\frac{dD_A^{(2)}}{d\tau} = -(\lambda^c + \lambda^c + \mu + \gamma) D_A^{(2)} + (\lambda^c E + \lambda^c E^2 + \mu) D_A^{(3)}$ $\frac{dD_A^{(3)}}{d\tau} = -\gamma D_A^{(3)} + \gamma D_M^{(3)}$ $\frac{dE}{d\tau} = \mu - (\lambda^c + \mu) E + \lambda^c E^2$	Initial conditions: $D_A^{(1)}(\tau_{\min}) = 0, D_A^{(2)}(\tau_{\min}) = 0, D_A^{(3)}(\tau_{\min}) = 1, E(\tau_{\min}) = E(\tau_{\min}).$ System of equations: $\frac{dD_A^{(1)}}{d\tau} = -(\lambda^c + \mu) D_A^{(1)} + 2\lambda^c D_A^{(1)} E$ $\frac{dD_A^{(2)}}{d\tau} = -(\lambda^c + \lambda^c + \mu + \gamma) D_A^{(2)} + (\lambda^c E + \lambda^c E^2 + \mu) D_A^{(3)}$ $\frac{dD_A^{(3)}}{d\tau} = -\gamma D_A^{(3)} + \gamma D_M^{(3)}$ $\frac{dE}{d\tau} = \mu - (\lambda^c + \mu) E + \lambda^c E^2$	Initial conditions: $D_A^{(1)}(\tau_{\max}) = 0, D_A^{(2)}(\tau_{\max}) = 0, D_A^{(3)}(\tau_{\max}) = 1, E(\tau_{\max}) = E(\tau_{\max}).$ System of equations: $\frac{dD_A^{(1)}}{d\tau} = -(\lambda^c + \mu) D_A^{(1)} + 2\lambda^c D_A^{(1)} E$ $\frac{dD_A^{(2)}}{d\tau} = -(\lambda^c + \lambda^c + \mu + \gamma) D_A^{(2)} + (\lambda^c E + \lambda^c E^2 + \mu) D_A^{(3)}$ $\frac{dD_A^{(3)}}{d\tau} = -\gamma D_A^{(3)} + \gamma D_M^{(3)}$ $\frac{dE}{d\tau} = \mu - (\lambda^c + \mu) E + \lambda^c E^2$	Initial conditions: $D_A^{(1)}(\tau_{\max}) = D_A^{(2)}(\tau_{\max}), D_M^{(1)}(\tau_{\max}) = D_M^{(2)}(\tau_{\max}), E(\tau_{\max}) = E(\tau_{\max}).$ System of equations: $\frac{dD_A^{(1)}}{d\tau} = -\gamma D_A^{(1)} + \gamma D_M^{(1)}$ $\frac{dD_M^{(1)}}{d\tau} = -(\lambda^c + \lambda^c + \mu) D_M^{(1)} + (\lambda^c E + \lambda^c E^2 + \mu) D_A^{(1)}$ $\frac{dE}{d\tau} = \mu - (\lambda^c + \mu) E + \lambda^c E^2$

- t_0, τ_0 : time of the island's origin, i.e. the island's age
- $t_{\max}, \tau_{\max}$: maximum time of colonization
- $t_{\min}, \tau_{\min}$: minimum time of colonization
- $t_{C,1}, \tau_{C,1}$: time of first colonization event in the colonization time window $[t_{\max}, t_{\min}]$

- t_C, τ_C : time of colonization event that gives rise the lineage observed at the present time
- t_E, τ_E : time at which the lineage at the present time becomes endemic
- t_p, τ_p : present day time

There are two possible orderings of these times:

- Upper panels in Figs. 4 b) and c): $t_C < t_{\min} < t_E$ (including the case $t_E = t_2$), if speciation of the observed lineages occurs after $t_{\min}$
- Lower panels in Figs. 4 b) and c): $t_C < t_E < t_{\min}$, if speciation of the observed lineages occurs between $[t_{\max}, t_{\min}]$

We construct the likelihoods by integrating a series of differential equations. To formulate these equations, it is convenient to introduce different instances of the variables D_A and D_M :

$$D_M^{(1)}(\tau) \quad \text{for } \tau \in [\tau_{c,1}, \tau_0] \quad (\text{A1a})$$

$$D_M^{(2)}(\tau) \quad \text{for } \tau \in [\tau_E, \tau_{c,1}] \quad (\text{A1b})$$

$$D_M^{(3)}(\tau) \quad \text{for } \tau \in [\tau_p, \tau_E] \quad (\text{A1c})$$

$$D_A^{(1)}(\tau) \quad \text{for } \tau \in [\tau_{\max}, \tau_0] \quad (\text{A1d})$$

$$D_A^{(2)}(\tau) \quad \text{for } \tau \in [\tau_C, \tau_{\max}] \quad (\text{A1e})$$

$$D_A^{(3)}(\tau) \quad \text{for } \tau \in [\tau_p, \tau_E] \quad (\text{A1f})$$

A.1 Singleton Endemic or Non Endemic Lineage with Colonization Time at τ_C

In this scenario, a mainland species colonizes the island at time τ_C . For endemic singleton lineages, speciation occurs at an unknown time $\tau_E \in [\tau_C, \tau_p]$. Before this event, the lineage is present on the island as a non-endemic species, and its dynamics are described by $D_M^{(2)}$. After speciation, and until the present time τ_p , the mainland ancestor may recolonize the island and subsequently go extinct again; this is captured by $D_M^{(3)}$ and $D_A^{(3)}$. For non-endemic singleton lineages, no speciation occurs, so the process remains in state $D_M^{(2)}$. The dynamical equations on $[\tau_p, \tau_C]$ are given by Eqs. (8a–8d). The equation is E are given by Eq. (2c)

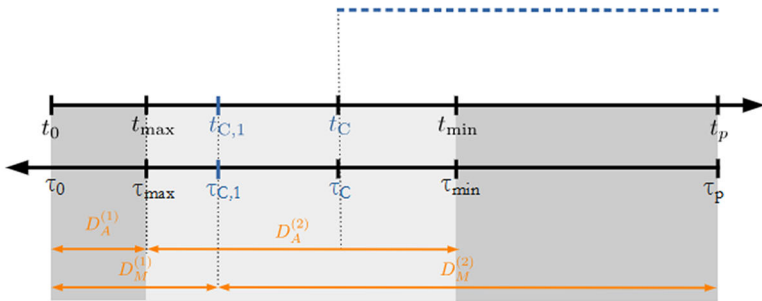
For endemic singleton lineages, the initial conditions at the present time τ_p are given by:

$$D_E(\tau_p) = 1, \quad D_M^{(2)}(\tau_p) = 0, \quad D_M^{(3)}(\tau_p) = 0, \quad D_A^{(3)}(\tau_p) = 1, \quad E(\tau_p) = 0. \quad (\text{A2})$$

These initial conditions reflect the fact that, at the present time τ_p , the lineage is observed as an endemic singleton.

For non-endemic singleton lineages, the lineage is always in the state $D_M^{(2)}(\tau)$ in the time interval $[\tau_p, \tau_C]$ because the mainland species never become endemic. The set of equation Eqs. (8a–8d) reduces to:

(a) Non-endemic singleton



(b) Endemic singleton

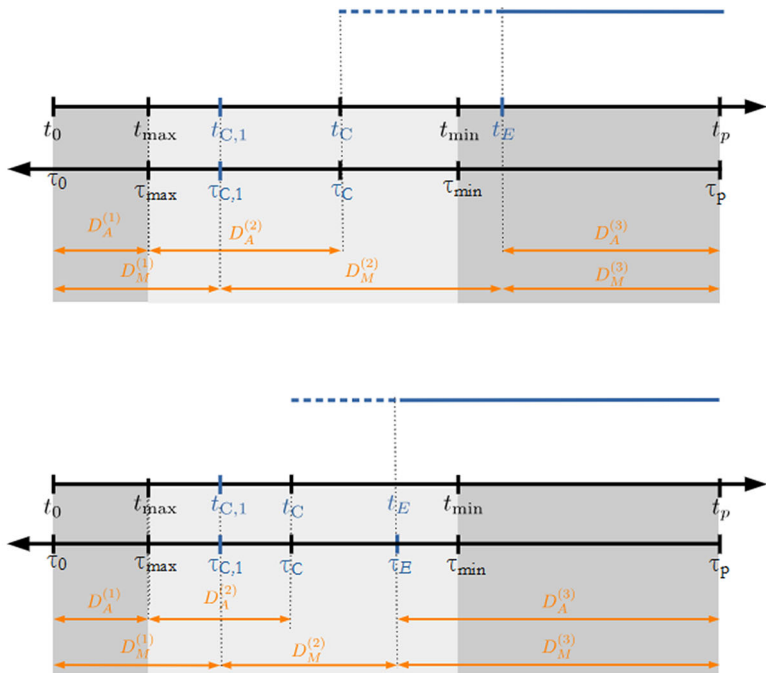


Fig. 4 Schematic representations of the temporal structure and definition of the state probabilities $D_M(\tau)$ and $D_A(\tau)$ in the backward pruning algorithm for different cases. The blue line represents the time interval during which the species is endemic. The orange arrows show the time intervals over which each probability is computed, going backward in time. The grey-shaded regions represent different time intervals, each governed by a distinct set of differential equations. These intervals correspond to different phases in the lineage history, reflecting changes in model dynamics across time

(c) Endemic radiation

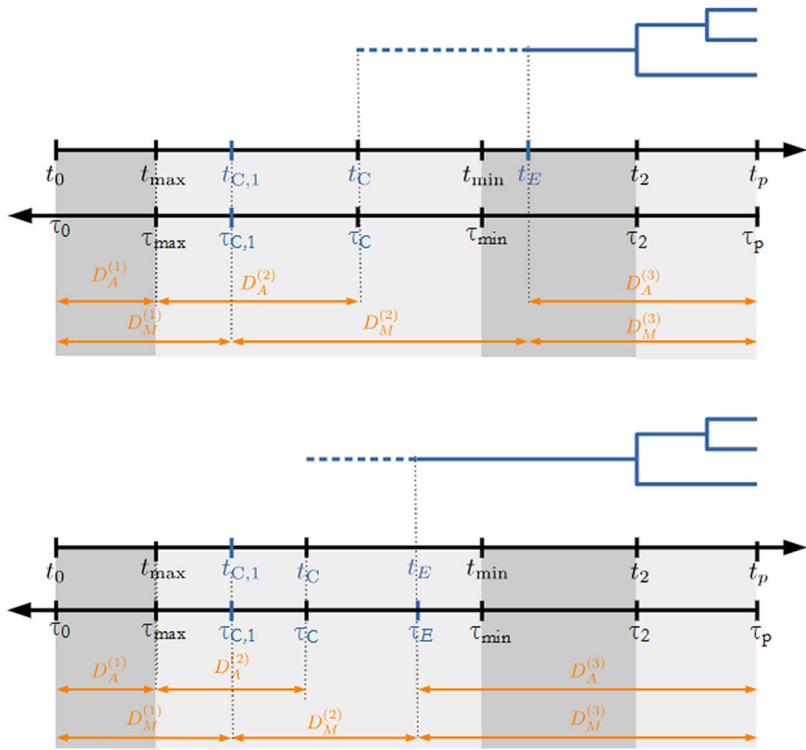


Fig. 4 continued

$$\frac{dD_M^{(2)}}{d\tau} = -(\lambda^a + \lambda^c + \mu + \gamma)D_M^{(2)} \quad (\text{no anagenesis/cladogenesis/extinction/colonization}) \quad (\text{A3a})$$

$$\begin{aligned} \frac{dE}{d\tau} &= \mu & (\text{extinction}) \\ &- (\lambda^c + \mu)E & (\text{no cladogenesis/extinction}) \\ &+ \lambda^c E^2 & (\text{cladogenesis followed by two extinctions}) \end{aligned} \quad (\text{A3b})$$

with initial conditions at τ_p given by:

$$D_M^{(2)}(\tau_p) = 1, \quad E(\tau_p) = 0. \quad (\text{A4})$$

In the interval $[\tau_C, \tau_0]$ there is no clade, and so the mainland species is not represented. Still it could have colonized the island, a possibility we describe with the variable $D_M^{(1)}$.

The dynamical equations describing the change in the probabilities between τ_C and τ_0 are given by Eqs. (2a–2c). The initial conditions at τ_C are given by:

$$D_A^{(1)}(\tau_C^+) = D_M^{(1)}(\tau_C^+) = \gamma D_M^{(2)}(\tau_C^-), \quad E(\tau_C^+) = E(\tau_C^-). \quad (A5)$$

A.2 Non-endemic Lineage with Maximum and Minimum Colonization Time

In this case, a non-endemic lineage is present on an island, with known maximum and minimum ages of colonization. In the interval $[\tau_p, \tau_{min}]$, the species is present and is non-endemic. The dynamical equations describing the change in the probabilities in this time interval are given by Eqs. (A3a–A3b). The observed species colonized the island between the maximum colonization time τ_{max} , and the minimum colonization time τ_{min} . If there is a colonization event in $[\tau_0, \tau_{max}]$ and the mainland species survives until τ_{max} , it should be replaced by a new colonization event after τ_{max} . To impose this, we keep track in the interval $[\tau_{min}, \tau_{max}]$ of whether the mainland species currently present originates from a colonization before or after τ_{max} . Hence, we need two variables $D_M^{(1)}$ and $D_M^{(2)}$. All colonization events in $[\tau_0, \tau_{max}]$ will be tracked using $D_M^{(1)}$. A new colonization after τ_{max} corresponds to a transition from $D_M^{(1)}$ (mainland species present, colonization before τ_{max}) to $D_M^{(2)}$ (mainland species present, colonization after τ_{max}). Therefore, in the interval $[\tau_{min}, \tau_{max}]$ the dynamical equations describing the change in the probabilities are given by:

$$\begin{aligned} \frac{dD_A^{(2)}}{d\tau} &= -\gamma D_A^{(2)} && \text{(no recolonization)} \\ &+ \gamma D_M^{(2)} && \text{(recolonization by mainland species)} \end{aligned} \quad (A6a)$$

$$\begin{aligned} \frac{dD_M^{(1)}}{d\tau} &= -(\lambda^a + \lambda^c + \mu + \gamma) D_M^{(1)} && \text{(no anagenesis/cladogenesis/extinction/colonization)} \\ &+ (\lambda^a E + \lambda^c E^2 + \mu) D_A^{(1)} && \text{(anagenesis followed by extinction; cladogenesis followed by two extinctions; extinction of mainland species; immediately after event mainland species is absent)} \\ &+ \gamma D_M^{(2)} && \text{(recolonization by mainland species)} \end{aligned} \quad (A6b)$$

$$\begin{aligned} \frac{dD_M^{(2)}}{d\tau} &= -(\lambda^a + \lambda^c + \mu + \gamma) D_M^{(2)} && \text{(no anagenesis/cladogenesis/extinction/colonisation)} \\ &+ (\lambda^a E + \lambda^c E^2 + \mu) D_A^{(2)} && \text{(anagenesis followed by extinction; cladogenesis followed by two extinctions; extinction of mainland species; immediately after event mainland species is absent)} \\ &+ \gamma D_M^{(2)} && \text{(recolonisation by mainland species)} \end{aligned} \quad (A6c)$$

$$\begin{aligned}
 \frac{dE}{d\tau} &= \mu && \text{(extinction)} \\
 &- (\lambda^c + \mu)E && \text{(no cladogenesis/extinction)} \\
 &+ \lambda^c E^2 && \text{(cladogenesis followed by extinction of both} \\
 &&& \text{daughter lineages)}
 \end{aligned} \tag{A6d}$$

with the initial conditions

$$D_A^{(2)}(\tau_{min}^+) = 0 \tag{A7a}$$

$$D_M^{(1)}(\tau_{min}^+) = 0 \tag{A7b}$$

$$D_M^{(2)}(\tau_{min}^+) = D_M^{(2)}(\tau_{min}^-) \tag{A7c}$$

$$E(\tau_{min}^+) = E(\tau_{min}^-). \tag{A7d}$$

In the interval $[\tau_{max}, \tau_0]$, the dynamical equations describing the change in the probabilities are given by Eqs. (2a–2c), with initial conditions

$$D_A^{(1)}(\tau_{max}^+) = D_A^{(2)}(\tau_{max}^-) \tag{A8a}$$

$$D_M^{(1)}(\tau_{max}^+) = D_M^{(1)}(\tau_{max}^-) \tag{A8b}$$

$$E(\tau_{max}^+) = E(\tau_{max}^-). \tag{A8c}$$

A.3 Endemic Singleton Lineage with Maximum and Minimum Colonization Time

In this case, an endemic singleton lineage is present on an island, with known maximum and minimum ages of colonization. If there is a colonization event in $[\tau_0, \tau_{max}]$ and the mainland species survives until τ_{max} , it should be replaced by a new colonization event after τ_{max} (in forward time). To impose this, we should keep track in the interval $[\tau_{max}, \tau_{min}]$ of whether the mainland species currently present originates from a colonization before or after τ_{max} . Hence, we need two variables D_M ($D_M^{(1)}$ and $D_M^{(2)}$). A new colonization after τ_{max} corresponds to a transition from $D_M^{(1)}$ (mainland species present, colonization before τ_{max}) to $D_M^{(2)}$ (mainland species present, colonization after τ_{max}). Between the speciation time and the present the mainland species can recolonize (and go extinct again before τ_p). This possibility is described by the variables $D_M^{(3)}$ and $D_A^{(3)}$.

In the interval $[\tau_p, \tau_{min}]$, the species is present, and may either still be non-endemic or have already undergone speciation. The dynamical equations describing the change in the probabilities in this time interval are given by:

$$\frac{dD_E}{d\tau} = -(\lambda^c + \mu)D_E \tag{no cladogenesis/extinction}$$

$$+ 2\lambda^c D_E E \quad (\text{cladogenesis followed by extinction of one daughter lineage}) \quad (\text{A9a})$$

$$\begin{aligned} \frac{dD_M^{(2)}}{d\tau} &= -(\lambda^a + \lambda^c + \mu + \gamma) D_M^{(2)} \quad (\text{no anagenesis/cladogenesis/extinction/colonization}) \\ &+ (\lambda^a D_E + 2\lambda^c D_E E) D_A^{(3)} \quad (\text{anagenesis; cladogenesis followed by extinction of one daughter lineage; immediately after event mainland species is absent}) \end{aligned} \quad (\text{A9b})$$

$$\begin{aligned} \frac{dD_M^{(3)}}{d\tau} &= -(\lambda^a + \lambda^c + \mu + \gamma) D_M^{(3)} \quad (\text{no anagenesis/cladogenesis/extinction/colonisation}) \\ &+ (\lambda^a E + \lambda^c E^2 + \mu) D_A^{(3)} \quad (\text{anagenesis followed by extinction; cladogenesis followed by two extinctions; extinction of mainland species; immediately after event mainland species is absent}) \\ &+ \gamma D_M^{(3)} \quad (\text{recolonisation by mainland species}) \end{aligned} \quad (\text{A9c})$$

$$\begin{aligned} \frac{dD_A^{(3)}}{d\tau} &= -\gamma D_A^{(3)} \quad (\text{no recolonization}) \\ &+ \gamma D_M^{(3)} \quad (\text{recolonization by mainland species}) \end{aligned} \quad (\text{A9d})$$

$$\begin{aligned} \frac{dE}{d\tau} &= \mu \quad (\text{extinction}) \\ &- (\lambda^c + \mu) E \quad (\text{no cladogenesis/extinction}) \\ &+ \lambda^c E^2 \quad (\text{cladogenesis followed by extinction of both daughter species}) \end{aligned} \quad (\text{A9e})$$

with the initial conditions at τ_p given by

$$D_E(\tau_p) = 1 \quad (\text{A10a})$$

$$D_M^{(2)}(\tau_p) = 0 \quad (\text{A10b})$$

$$D_M^{(3)}(\tau_p) = 0 \quad (\text{A10c})$$

$$D_A^{(3)}(\tau_p) = 1 \quad (\text{A10d})$$

$$E(\tau_p) = 0. \quad (\text{A10e})$$

For the interval $[\tau_{min}, \tau_{max}]$, the species has colonized and may or may not have undergone speciation. The dynamical equations describing the change in the probabilities in this time interval are given by:

$$\frac{dD_E}{d\tau} = -(\lambda^c + \mu) D_E \quad (\text{no cladogenesis/extinction})$$

$$+ 2\lambda^c D_E E \quad (\text{cladogenesis followed by extinction of one daughter lineage}) \quad (\text{A11a})$$

$$\frac{dD_A^{(2)}}{d\tau} = -\gamma D_A^{(2)} \quad (\text{no recolonization})$$

$$+ \gamma D_M^{(2)} \quad (\text{recolonization by mainland species}) \quad (\text{A11b})$$

$$\frac{dD_A^{(3)}}{d\tau} = -\gamma D_A^{(3)} \quad (\text{no recolonization})$$

$$+ \gamma D_M^{(3)} \quad (\text{recolonization by mainland species}) \quad (\text{A11c})$$

$$\frac{dD_M^{(1)}}{d\tau} = -(\lambda^a + \lambda^c + \mu + \gamma) D_M^{(1)} \quad (\text{no anagenesis/cladogenesis/extinction/colonization})$$

$$+ (\lambda^a E + \lambda^c E^2 + \mu) D_A^{(1)} \quad (\text{anagenesis followed by extinction; cladogenesis followed by two extinctions; extinction of mainland species; immediately after event mainland species is absent})$$

$$+ \gamma D_M^{(2)} \quad (\text{recolonization by mainland species}) \quad (\text{A11d})$$

$$\frac{dD_M^{(2)}}{d\tau} = -(\lambda^a + \lambda^c + \mu + \gamma) D_M^{(2)} \quad (\text{no anagenesis/cladogenesis/extinction/colonisation})$$

$$+ (\lambda^a D_E + 2\lambda^c D_E E) D_A^{(3)} \quad (\text{anagenesis; cladogenesis followed by extinction of one daughter lineage; immediately after event mainland species is absent and recolonization does not leave descendants at present})$$

$$+ (\lambda^a E + \lambda^c E^2 + \mu) D_A^{(2)} \quad (\text{anagenesis; cladogenesis followed by extinction of one daughter lineage; immediately after event mainland species is absent and recolonization can initiate observed lineage})$$

$$+ \gamma D_M^{(2)} \quad (\text{recolonisation by mainland species}) \quad (\text{A11e})$$

$$\frac{dD_M^{(3)}}{d\tau} = -(\lambda^a + \lambda^c + \mu + \gamma) D_M^{(3)} \quad (\text{no anagenesis/cladogenesis/extinction/colonisation})$$

$$+ (\lambda^a E + \lambda^c E^2 + \mu) D_A^{(3)} \quad (\text{anagenesis followed by extinction; cladogenesis followed by two extinctions; extinction of mainland species; immediately after event mainland species is absent})$$

$$+ \gamma D_M^{(3)} \quad (\text{recolonisation by mainland species}) \quad (\text{A11f})$$

$$\frac{dE}{d\tau} = \mu \quad (\text{extinction})$$

$$- (\lambda^c + \mu) E \quad (\text{no cladogenesis/extinction})$$

$$+ \lambda^c E^2 \quad (\text{cladogenesis followed by extinction of both daughter species}) \quad (\text{A11g})$$

with the initial conditions

$$D_E(\tau_{min}^+) = D_E(\tau_{min}^-), \quad (\text{A12a})$$

$$D_A^{(3)}(\tau_{min}^+) = D_A^{(3)}(\tau_{min}^-), \quad (\text{A12b})$$

$$D_A^{(2)}(\tau_{min}^+) = 0, \quad (\text{A12c})$$

$$D_M^{(1)}(\tau_{min}^+) = 0, \quad (\text{A12d})$$

$$D_M^{(3)}(\tau_{min}^+) = D_M^{(3)}(\tau_{min}^-), \quad (\text{A12e})$$

$$D_M^{(2)}(\tau_{min}^+) = D_M^{(2)}(\tau_{min}^-), \quad (\text{A12f})$$

$$E(\tau_{min}^+) = E(\tau_{min}^-). \quad (\text{A12g})$$

For the interval $[\tau_{max}, \tau_0]$, the dynamical equations are given by Eqs. (2a–2c), with initial conditions

$$D_A^{(1)}(\tau_{max}^+) = D_A^{(2)}(\tau_{max}^-) \quad (\text{A13a})$$

$$D_M^{(1)}(\tau_{max}^+) = D_M^{(1)}(\tau_{max}^-) \quad (\text{A13b})$$

$$E(\tau_{max}^+) = E(\tau_{max}^-). \quad (\text{A13c})$$

If the endemic singleton species coexists on the island with its mainland ancestors, the initial conditions of the set of equations Eqs. (A9a–A9e) are modified to

$$D_E(\tau_p) = 1 \quad (\text{A14a})$$

$$D_M^{(2)}(\tau_p) = 0 \quad (\text{A14b})$$

$$D_M^{(3)}(\tau_p) = 1 \quad (\text{A14c})$$

$$D_A^{(2)}(\tau_p) = 0 \quad (\text{A14d})$$

$$D_A^{(3)}(\tau_p) = 0 \quad (\text{A14e})$$

$$E(\tau_p) = 0. \quad (\text{A14f})$$

A.4 Endemic Non-singleton Lineage with Maximum Colonization Time

In this case, an endemic lineage is present on an island, for which only the maximum age of colonization is known.

In the interval $[\tau_2, \tau_p]$, the species has already undergone cladogenesis speciation. For the branches represented in the clade, the dynamical equations describing the change in the probabilities in this time interval are given by Eqs. (5a–5d), with the initial

conditions

$$D_E(\tau_p) = 1 \quad (\text{A15a})$$

$$D_M^{(3)}(\tau_p) = 0 \quad (\text{A15b})$$

$$D_A^{(3)}(\tau_p) = 1 \quad (\text{A15c})$$

$$E(\tau_p) = 0. \quad (\text{A15d})$$

This set of equations is integrated along the edges of the trees starting from the tips. At each branching time we have to stop the integration and set $D_E(\tau_i)$ to $\lambda^c d\tau$ times the product of the D_E functions of the merging subclades (and do not change the E function).

In the time interval $[\tau_{max}, \tau_2]$ we have to add the dynamical equation for $D_M^{(2)}$, because we know that on at least a part of the branch the species is non-endemic. The dynamical equations describing the change in the probabilities in this time interval are given by eqs. (A11a–A11g). As initial conditions for $D_E(\tau_2^+)$ we take λ^c (from $\lambda^c d\tau$ which represents the probability of cladogenesis speciation over the small time interval $d\tau$) times the product of the D_E functions of the merging subclades. For $D_M^{(2)}(\tau_2^+)$ we use the same quantity multiplied by the probability $D_A(\tau_2^-)$ that the mainland species is not present on the island. This covers both cases in which the speciation event that transforms the mainland species into an endemic one occurs either between τ_{max} and τ_2 or precisely at time τ_2 . The initial conditions are then given by:

$$D_E(\tau_2^+) = \lambda^c D_E^{(l)}(\tau_2^-) D_E^{(r)}(\tau_2^-) \quad (\text{A16a})$$

$$D_M^{(1)}(\tau_2^+) = 0 \quad (\text{A16b})$$

$$D_M^{(2)}(\tau_2^+) = \lambda^c D_E^{(l)}(\tau_2^-) D_E^{(r)}(\tau_2^-) D_A(\tau_2^-) \quad (\text{A16c})$$

$$D_M^{(3)}(\tau_2^+) = D_M^{(3)}(\tau_2^-) \quad (\text{A16d})$$

$$D_A^{(2)}(\tau_2^+) = 0 \quad (\text{A16e})$$

$$D_A^{(3)}(\tau_2^+) = D_A^{(3)}(\tau_2^-) \quad (\text{A16f})$$

$$E(\tau_2^+) = E(\tau_2^-) \quad (\text{A16g})$$

If species within lineages are sampled at random (all species in the clade have the same probability of being sampled), incomplete sampling can be accounted for by modifying the initial conditions for each tip in the tree to reflect ρ , the probability of sampling a species (following FitzJohn et al. (2009)). In this case, the initial conditions become $D_E(\tau_p) = \rho$ and $E(\tau_p) = 1 - \rho$.

In the interval $[\tau_{max}, \tau_0]$, the dynamical equations describing the change in the probabilities are given by Eqs. (2a–2c) with initial conditions

$$D_A^{(1)}(\tau_{max}^+) = D_A^{(2)}(\tau_{max}^-), \quad (\text{A17a})$$

$$D_M^{(1)}(\tau_{max}^+) = D_M^{(1)}(\tau_{max}^-), \quad (\text{A17b})$$

$$E(\tau_{max}^+) = E(\tau_{max}^-). \quad (\text{A17c})$$

If the endemic singleton species coexists on the island with its mainland ancestors, the initial conditions of the set of equations Eqs. (5a–5d) are modified to:

$$D_E(\tau_p) = 1, \quad (\text{A18a})$$

$$E(\tau_p) = 0, \quad (\text{A18b})$$

$$D_A^{(3)}(\tau_p) = 0, \quad (\text{A18c})$$

$$D_M^{(3)}(\tau_p) = 1. \quad (\text{A18d})$$

Appendix B Analytical Likelihood Expressions for Special Cases

Here we present analytical likelihood expressions for a series of special cases in order to provide additional support to both the DAISIE and DAISIE-DE models. To facilitate the derivations, we assume that both extinction and anagenesis are absent, i.e., we set $\mu = \lambda^a = 0$. Let t_0 denote the island age, t_C the colonization time, t_2 the branching time, and t_p the present time. All times are expressed in forward time, so that $t_0 < t_C < t_2 < t_p$.

B.1 Dynamics Before Colonization Time

Before the colonization event of the observed lineage, the mainland species may have previously colonized the island. If this occurred, the earlier colonizing lineage must not have undergone speciation, as its descendant lineages cannot go extinct under the model assumptions. If a mainland species colonized the island prior to the estimated colonization time of the observed lineage and survive until that time, its dynamics are assumed to be taken over by that of the newly colonizing lineage, in accordance with the DAISIE and DAISIE-DE model assumptions. This gives the following dynamics before the colonization time:

$$\frac{dP_0}{dt} = -\gamma P_0(t) \quad (\text{B1a})$$

$$\frac{dP_1}{dt} = \gamma P_0(t) - \lambda^c P_1(t) \quad (\text{B1b})$$

with solutions:

$$P_0(t; p) = e^{-\gamma(t-t_0)}(1 - p(t)) \quad (\text{B2a})$$

$$P_1(t; p) = \frac{\gamma}{\lambda^c - \gamma} \left(e^{-\gamma(t-t_0)} - e^{-\lambda^c(t-t_0)} \right) (1 - p(t)) + e^{-\lambda^c(t-t_0)} p(t) \quad (\text{B2b})$$

where p is the probability that the mainland species is present at time t_0 . The likelihood is given by

$$L_0 = P_0(t; p) + P_1(t; p) \quad (\text{B3})$$

B.2 Endemic Clade with Two Tips, Known Colonization Time

For an endemic clade that colonizes the island at time t_C and speciated at time t_2 , the likelihood formula is:

$$L_1 = (P_0(t_C; 0) + P_1(t_C; 0)) \gamma e^{-(\lambda^c + \gamma)(t_2 - t_C)} \lambda^c e^{-(2\lambda^c + \gamma)(t_p - t_2)} \quad (\text{B4a})$$

$$= \left(\frac{\lambda^c}{\lambda^c - \gamma} e^{-\gamma(t_C - t_0)} - \frac{\gamma}{\lambda^c - \gamma} e^{-\lambda^c(t_C - t_0)} \right) \gamma e^{-(\lambda^c + \gamma)(t_2 - t_C)} \lambda^c e^{-(2\lambda^c + \gamma)(t_p - t_2)} \quad (\text{B4b})$$

The expression $(P_0(t_C; 0) + P_1(t_C; 0))$ describes the probability dynamics before the colonization event at time t_C . The exponential term $e^{-(\lambda^c + \gamma)(t_2 - t_C)}$ represents the probability of no events (neither cladogenesis, nor extinction) occurring between the colonization time t_C and the cladogenesis event at t_2 . Finally, the term $e^{-(2\lambda^c + \gamma)(t_p - t_2)}$ accounts for the probability that both daughter lineages, resulting from cladogenesis at t_2 survive until the present time t_p without further diversification and colonization.

B.3 Endemic Clade with Two Tips, Unknown Colonization Time

The likelihood formula is:

$$L_2 = P_1(t_2; 0) \lambda^c e^{-(2\lambda^c + \gamma)(t_p - t_2)} \quad (\text{B5a})$$

$$= \frac{\gamma}{\lambda^c - \gamma} \left(e^{-\gamma(t_2 - t_0)} - e^{-\lambda^c(t_2 - t_0)} \right) \lambda^c e^{-(2\lambda^c + \gamma)(t_p - t_2)} \quad (\text{B5b})$$

$P_1(t_2; 0)$ describes the probability of observing a singleton lineage on the island, that have colonized at a time t between the island age t_0 , and the cladogenesis time t_2 , the term $e^{-(2\lambda^c + \gamma)(t_p - t_2)}$ accounts for the probability that both daughter lineages, resulting from cladogenesis at t_2 survive until the present time t_p without further diversification and colonization. It can be checked that:

$$L_2 = \int_{t_0}^{t_2} L_1 dt_C \quad (\text{B6})$$

B.4 Endemic Clade with Two Tips, Maximum Colonization Time

If there is a maximum colonization time t_{m1} , and the branching time t_2 is the minimum colonization time, we subtract from the probability that a colonization takes place before t_2 ($P_1(t_2; 0)$) the probability that a colonization takes place before t_{m1} without a new colonization between t_{m1} and t_2 . The likelihood formula is:

$$L_3 = \left(P_1(t_2; 0) - P_1(t_{m1}; 0) e^{-(\lambda^c + \gamma)(t_2 - t_{m1})} \right) \lambda^c e^{-(2\lambda^c + \gamma)(t_p - t_2)} \quad (\text{B7a})$$

$$= \frac{\gamma}{\lambda^c - \gamma} \left(e^{-\gamma(t_2 - t_0)} - e^{-\lambda^c(t_2 - t_0)} - e^{-\gamma(t_{m1} - t_0)} e^{-(\lambda^c + \gamma)(t_2 - t_{m1})} + e^{-\lambda^c(t_{m1} - t_0)} e^{-(\lambda^c + \gamma)(t_2 - t_{m1})} \right) \lambda^c e^{-(2\lambda^c + \gamma)(t_p - t_2)} \quad (\text{B7b})$$

$(P_1(t_2; 0) - P_1(t_{m1}; 0)e^{-(\lambda^c + \gamma)(t_2 - t_{m1})})$ describes the probability of observing a singleton lineage on the island, that have colonized between the time t_{m1} and the cladogenesis time t_2 . The term $e^{-(2\lambda^c + \gamma)(t_p - t_2)}$ accounts for the probability that both daughter lineages, resulting from cladogenesis at t_2 survive until the present time t_p without further diversification and colonization. L_2 can be recovered from L_3 by setting $t_{m1} = t_0$.

B.5 Endemic Clade with Two Tips that Colonized at Time t_0

If we know that at time t_0 the island is already colonized by the mainland species, the likelihood would be:

$$P_1(t_2; 1)\lambda^c e^{-(2\lambda^c + \gamma)(t_p - t_2)} = e^{-\lambda^c(t_2 - t_0)}\lambda^c e^{-(2\lambda^c + \gamma)(t_p - t_2)} \quad (\text{B8})$$

If the mainland species is present on the island with probability p , the likelihood is:

$$\begin{aligned} L_4 &= P_1(t_2; p)\lambda^c e^{-(2\lambda^c + \gamma)(t_p - t_2)} \\ &= (1 - p(t))\frac{\gamma}{\lambda^c - \gamma} \left(e^{-\gamma(t_2 - t_0)} - e^{-\lambda^c(t_2 - t_0)} \right) \lambda^c e^{-(2\lambda^c + \gamma)(t_p - t_2)} \\ &\quad + p(t) e^{-\lambda^c(t_2 - t_0)}\lambda^c e^{-(2\lambda^c + \gamma)(t_p - t_2)} \end{aligned} \quad (\text{B9})$$

Here, we assume that if the mainland species is not present on the island at time t_0 , it can immigrate immediately afterwards.

Appendix C Relating ρ -sampling to n -sampling

In our approach we use the sampling fraction ρ to specify the incompleteness of an empirical phylogeny. This is referred to as ρ -sampling. By contrast, DAISIE requires the number of unsampled species to be specified, which we call n -sampling. For comparison of the results with our approach with DAISIE, we would like to compute the likelihood under n -sampling from the likelihood under ρ -sampling. The likelihoods of phylogenetic data (which we will refer to as ‘clade’) under the two types of sampling are related by:

$$P_\rho(\text{clade}) = \sum_{n=0}^{\infty} P_n(\text{clade}) \binom{S+n}{n} \rho^S (1-\rho)^n \quad (\text{C1})$$

where S is the number of species in the clade, n the number of missing species, and ρ the probability of a species to be sampled and included as assumed in the ρ -sampling. The goal here is to isolate P_n from Eq. (1). To do so, we divide by ρ^S on both sides:

$$\frac{P_\rho(\text{clade})}{\rho^S} = \sum_{n=0}^{\infty} P_n(\text{clade}) \binom{S+n}{n} (1-\rho)^n \quad (\text{C2})$$

Let's now define $x = 1 - \rho$ and write the equation in terms of x :

$$\frac{P_{1-x}(\text{clade})}{(1-x)^S} = \sum_{n=0}^{\infty} P_n(\text{clade}) \binom{S+n}{n} x^n \quad (\text{C3})$$

The right-hand side is a polynomial in x , which we identify term by term with the Taylor expansion of the left-hand side about $x = 0$. This yields

$$P_n(\text{clade}) \binom{S+n}{n} = \frac{1}{n!} \left. \frac{d^n f(x)}{dx^n} \right|_{x=0} \quad (\text{C4})$$

and hence

$$P_n(\text{clade}) = \frac{S!}{(S+n)!} \left. \frac{d^n f(x)}{dx^n} \right|_{x=0} \quad (\text{C5})$$

with

$$f(x) = \frac{P_{1-x}(\text{clade})}{(1-x)^S} \quad (\text{C6})$$

What remains is to find a numerically stable way to determine the derivatives. Because we actually calculate the logarithm of $P_{1-x}(\text{clade})$, it is convenient to write the n -th order derivative in terms of the derivatives of $g(x) = \log f(x)$:

$$f^{(n)}(x) = f(x) B_n(g'(x), g''(x), \dots, g^{(n)}(x)) \quad (\text{C7})$$

Here,

$$g(x) = \log(f(x)) = \log(P_{1-x}(\text{clade})) - S \log(1-x) \quad (\text{C8})$$

and B_n are the n -th order Bell polynomials, given by

$$B_0 = 1, \quad B_n = \sum_{k=1}^n \binom{n-1}{k-1} B_{n-k} g^{(k)}(x). \quad (\text{C9})$$

which can be computed recursively. This is currently implemented in DAISIE.

As an alternative to the numerical evaluation of the derivatives, one can use complex function theory, and write the derivatives as contour integrals using Cauchy's integral formula. We then have

$$P_n(\text{clade}) = \frac{1}{\binom{S+n}{n}} \frac{1}{2\pi i} \oint \frac{f(z)}{z^{n+1}} dz \quad (\text{C10})$$

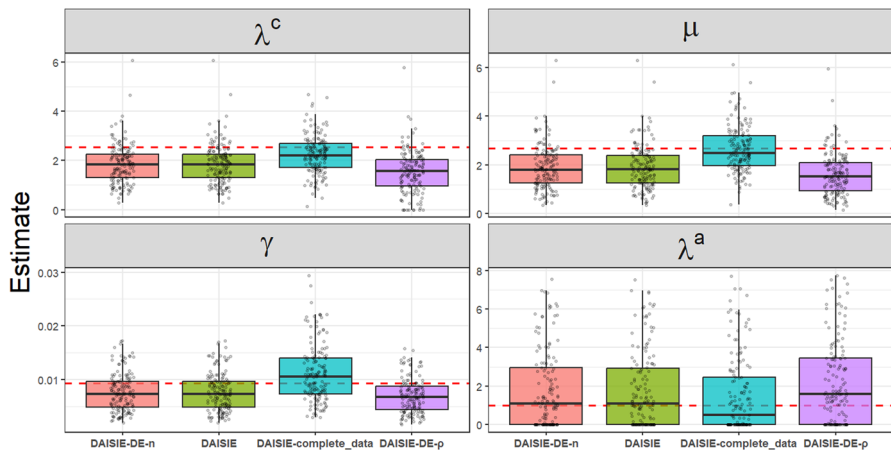


Fig. 5 Distribution of parameter estimates across models and 100 complete and incomplete replicates. Boxplots show the variability of the maximum-likelihood estimates for each parameter across simulation replicates and fitted models. The red dashed horizontal line indicates the true parameter value used to generate the simulated data. Points correspond to individual replicates. The complete-data model is shown as a reference, while the other methods are fitted to incomplete datasets with 50% data coverage

where $f(z)$ is the analytical continuation of $f(x)$ defined in Eq. (C6) to the complex plane:

$$f(z) = \frac{P_{1-z}(\text{clade})}{(1-z)^S} \quad (\text{C11})$$

and the contour goes around $z = 0$ in the positive direction (e.g. the unit circle). This approach only works if we can evaluate $P_\rho(\text{clade})$ for complex ρ . The ODE solvers currently implemented in DAISIE do not allow this, so this will be an avenue to explore in the future.

Finally, we note that because the ODE system is linear in D_E , we can replace the initial condition $D_E = \rho$ by $D_E = 1$ when we at the same time replace $f(z)$ ($f(x)$) by $P_{1-z}(\text{clade})$ ($P_{1-x}(\text{clade})$). Figs. 5, 6 and 7 show the comparison of the ML-estimate obtained by fitting 100 generated incomplete datasets with 50% data coverage to the DAISIE model that uses the n -sampling approach, and the DAISIE-DE with both the n - and the ρ -sampling methods. We see that the parameter estimates obtained with DAISIE-DE using n -sampling and with the DAISIE model are practically identical, except for the anagenesis rate. This difference may be attributed to the difficulty of estimating the anagenesis rate accurately.

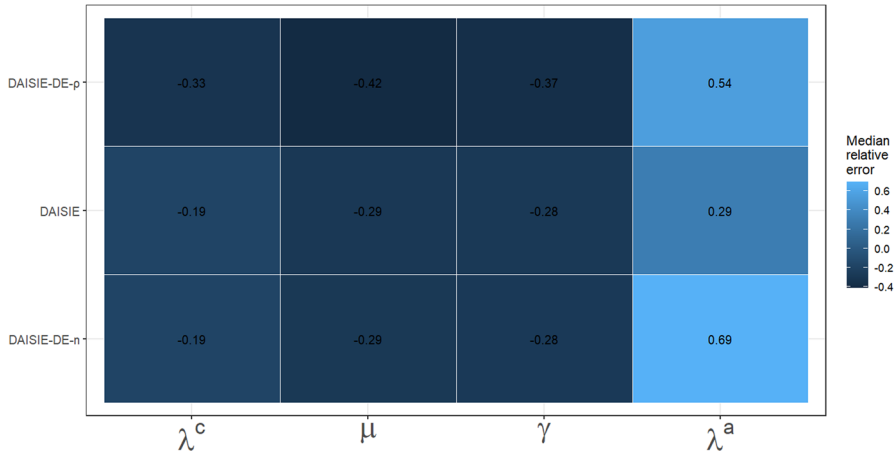


Fig. 6 Median relative deviation of parameter estimates from the complete-data reference model. Each cell shows the median relative error of a given method compared with the estimates obtained by fitting the model to the incomplete simulation replicates with 50% data coverage. Rows correspond to the models fitted to incomplete data, and columns correspond to the four estimated parameters λ^c —cladogenesis, μ —extinction, γ —colonization, λ^a —anagenesis rates (color figure online)

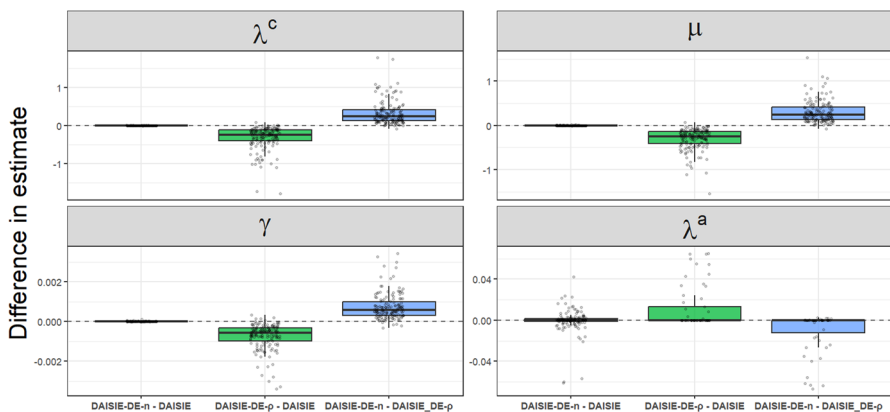


Fig. 7 Pairwise differences between parameter estimates obtained by fitting models to incomplete data with 50% data coverage. For each parameter, boxplots show the distribution of the difference between estimates obtained with pairs of models across incomplete replicates. Deviations from zero indicate differences induced by the choice of missing-data correction (color figure online)

Acknowledgements We would like to thank Ryan Brewer, Frederic Lens, Thijs Janzen and Luis Valente for their input, the Center for Information Technology of the University of Groningen for their support and for providing access to the Peregrine and Hábrók high-performance computing clusters, and the Dutch Research Council for financial support through a KLEIN grant awarded to Frederic Lens and Rampal S. Etienne (OCENW.KLEIN.498).

Declarations

Conflict of interest The authors declare no potential conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Caujapé-Castells J, García-Verdugo C, Marrero-Rodríguez Á, Fernández-Palacios JM, Crawford DJ, Mort ME (2017) Island ontogenies, syngameons, and the origins and evolution of genetic diversity in the canarian endemic flora. *Perspect Plant Ecol Evolut Syst* 27:9–22. <https://doi.org/10.1016/j.ppees.2017.03.003>
- Condamine FL, Rolland J, Morlon H (2013) Macroevolutionary perspectives to environmental change. *Ecol Lett* 16:72–85
- Etienne RS, Haegeman B, Dugo-Cota Á, Vilà C, Gonzalez-Voyer A, Valente L (2023) The phylogenetic limits to diversity-dependent diversification. *Syst Biol* 72(2):433–445. <https://doi.org/10.1093/sysbio/syac074>
- Etienne RS, Haegeman B, Stadler T, Aze T, Pearson PN, Purvis A, Phillimore AB (2012) Diversity-dependence brings molecular phylogenies closer to agreement with the fossil record. *Proc R Soc B Biol Sci* 279(1732):1300–1309. <https://doi.org/10.1098/rspb.2011.1439>
- Etienne RS, Rosindell J (2012) Prolonging the past counteracts the pull of the present: protracted speciation can explain observed slowdowns in diversification. *Syst Biol* 61(2):204. <https://doi.org/10.1093/sysbio/syr091>
- FitzJohn RG, Maddison WP, Otto SP (2009) Estimating trait-dependent speciation and extinction rates from incompletely resolved phylogenies. *Syst Biol* 58(6):595–611. <https://doi.org/10.1093/sysbio/syp067>
- Gillespie RG, Roderick GK (2014) Geology and climate drive diversification. *Nature* 509(7500):297–298
- Hauffe T, Delicado D, Etienne RS, Valente L (2020) Lake expansion elevates equilibrium diversity via increasing colonization. *J Biogeogr* 47(9):1849–1860. <https://doi.org/10.1111/jbi.13914>
- Helmus MR, Mahler DL, Losos JB (2014) Island biogeography of the anthropocene. *Nature* 513(7519):543–546
- Lambert JW, Neves PS, Bilderbeek RJ, Valente L, Etienne RS (2022) The effect of mainland dynamics on data and parameter estimates in island biogeography. <https://doi.org/10.1101/2022.01.13.476210>
- Losos JB, Ricklefs RE (2009) Adaptation and diversification on islands. *Nature* 457(7231):830–836
- MacArthur RH, Wilson EO (2001) *The theory of island biogeography*, vol 1. Princeton University Press
- Maddison WP (2006) Confounding asymmetries in evolutionary diversification and character change. *Evolution* 60(8):1743–1746. <https://doi.org/10.1111/j.0014-3820.2006.tb00517.x>
- Maddison WP, Midford PE, Otto SP (2007) Estimating a binary character's effect on speciation and extinction. *Syst Biol* 56(5):701–710. <https://doi.org/10.1080/10635150701607033>
- Morlon H (2014) Phylogenetic approaches for studying diversification. *Ecol Lett* 17(4):508–525. <https://doi.org/10.1111/ele.12251>
- Morlon H, Robin S, Hartig F (2022) Studying speciation and extinction dynamics from phylogenies: addressing identifiability issues. *Trend Ecol Evolut* 37(6):497–506
- Nee S, May RM, Harvey PH (1994) The reconstructed evolutionary process. *Philos Trans R Soc Lond Series B Biol Sci* 344(1309):305–311
- Neves PS, Lambert JW, Valente L, Etienne RS (2021) The robustness of a simple dynamic model of island biodiversity to geological and sea-level change. *J Biogeogr* 49(11):2091–2104. <https://doi.org/10.1111/jbi.14519>
- O'Connor MI, Mori AS, Gonzalez A, Dee LE, Loreau M, Avolio M, Byrnes JE, Cheung W, Cowles J, Clark AT et al (2021) Grand challenges in biodiversity-ecosystem functioning research in the era of science-policy platforms require explicit consideration of feedbacks. *Proc R Soc B* 288(1960):20210783. <https://doi.org/10.1098/rspb.2021.0783>

- Patino J, Whittaker RJ, Borges PA, Fernández-Palacios JM, Ah-Peng C, Araújo MB, Ávila SP, Cardoso P, Cornuault J, de Boer EJ et al (2017) A roadmap for island biology: 50 fundamental questions after 50 years of the theory of island biogeography. *J Biogeogr* 44(5):963–983. <https://doi.org/10.1111/jbi.12986>
- Valente L, Etienne RS, Davalos LM (2017) Recent extinctions disturb path to equilibrium diversity in caribbean bats. *Nat Ecol Evolut* 1(2):1–7
- Valente L, Etienne RS, Garcia-R JC (2019) Deep macroevolutionary impact of humans on new zealand's unique avifauna. *Curr Biol* 29(15):2563–2569
- Valente L, Illera JC, Havenstein K, Pallien T, Etienne RS, Tiedemann R (2017) Equilibrium bird species diversity in atlantic islands. *Curr Biol* 27(11):1660–1666
- Valente L, Phillimore AB, Etienne R (2018) Using molecular phylogenies in island biogeography: It's about time. *Ecography* 41(10):1684–1686
- Valente L, Phillimore AB, Melo M, Warren BH, Clegg SM, Havenstein K, Tiedemann R, Illera JC, Thébaud C, Aschenbach T et al (2020) A simple dynamic model explains the diversity of island birds worldwide. *Nature* 579(7797):92–96
- Valente LM, Phillimore AB, Etienne RS (2015) Equilibrium and non-equilibrium dynamics simultaneously operate in the galápagos islands. *Ecol Lett* 18(8):844–852. <https://doi.org/10.1111/ele.12461>
- Warren BH, Simberloff D, Ricklefs RE, Aguilée R, Condamine FL, Gravel D, Morlon H, Mouquet N, Rosindell J, Casquet J et al (2015) Islands as model systems in ecology and evolution: prospects fifty years after macarthur-wilson. *Ecol Lett* 18(2):200–217. <https://doi.org/10.1111/ele.12398>
- Xie S, Valente L, Etienne RS (2023) Can we ignore trait-dependent colonization and diversification in island biogeography? *Evolution* 77(3):670–681. <https://doi.org/10.1093/evolut/qpad006>
- Yadav D, Ishan Y, Pandya E, Manish D, Marjadi K, Swapnavahini, Kishore G, Lanka S, Swarupa M (2023). An Overview on Biodiversity: Life Forms of Nature, pp. 100–108. Bharti Publications

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.