

# DAISIEprep: an R package for the extraction and formatting of data for the island biogeography model DAISIE

Joshua W. Lambert<sup>1\*</sup>, Lizzie Roeble<sup>1,2</sup>, Théo Pannetier<sup>1</sup>,  
Rampal S. Etienne<sup>a,1</sup>, Luis Valente<sup>a,1,2</sup>

<sup>1</sup> *Groningen Institute for Evolutionary Life Sciences, University of Groningen, Box 11103, 9700 CC Groningen, The Netherlands*

<sup>2</sup> *Naturalis Biodiversity Center, Darwinweg 2, 2333 CR Leiden, The Netherlands*

\*Corresponding Author: Joshua W. Lambert, Groningen Institute for Evolutionary Life Sciences, Box 11103, 9700 CC Groningen, The Netherlands. Email: [j.w.l.lambert@rug.nl](mailto:j.w.l.lambert@rug.nl).

<sup>a</sup> These authors contributed jointly to senior authorship.

## Abstract

1. Phylogenetic trees are commonly used to answer questions on biogeographical and diversification histories of different groups.
2. Recently, new approaches have been developed that use community phylogenetic trees requiring a data structure distinct from the single phylogenetic trees that are commonly used, which may be a barrier to the utilisation of these approaches.
3. DAISIE (Dynamic Assembly of Islands through Speciation, Immigration and Extinction) is an island biogeography model that can estimate rates of colonisation, speciation and extinction from phylogenetic data across insular communities, as well as simulate islands under those rates.
4. Here we describe the **DAISIEprep** R package, a set of pre-processing tools to aid the extraction of data from one or many phylogenetic trees to generate data in a format interpretable by DAISIE for the application of island biogeography inference models. We present examples to illustrate the various data types that can be used.
5. The package includes simple algorithms to extract data on island colonists and account for biogeographical, topological and taxonomic uncertainty. It also allows flexible incorporation of either missing species or entire insular lineages when phylogenetic data are not available.
6. **DAISIEprep** enables reproducible and user-friendly data extraction and formatting, and will facilitate addressing questions about island biogeography, diversification and anthropogenic impacts in insular systems.

**Keywords:** community assembly, DAISIE, data science, island biogeography, macroevolution, phylogenetics, R package.

# Introduction

Phylogenetic trees are prevalent in evolutionary, epidemiological, and even linguistic sciences, and typically use well-defined data structures. Standard tree data structures, particularly the commonly used ‘Newick’ format (Maddison *et al.* 1997; Felsenstein 2004; Cardona *et al.* 2008), are employed within and across programming languages. For example, in the R language, phylogenetic tree formats have been constructed that are compatible with different packages and methodologies (Paradis *et al.* 2004; FitzJohn 2012; Revell 2012; Pennell *et al.* 2014; Morlon *et al.* 2016; R Core Team 2022). However, newly developed methods that use phylogenetic information may require different data structures, potentially creating obstacles for the scientific community to apply those methods. In particular, island biogeography research is increasingly utilising phylogenetic information to answer questions on island community assembly and diversification dynamics, but it requires phylogenetic information in an unconventional format. Instead of relying on a single phylogenetic tree representing a clade of interest, it can comprise data assembled from multiple phylogenetic trees of different clades and even non-phylogenetic information. In contrast to the phylogenetic tree formats, standards and methods for manipulating island community data are lacking.

DAISIE (Dynamic Assembly of Island biota through Speciation, Immigration and Extinction) is a model of island biogeography and macroevolution that focuses on entire island communities or assemblages, complementing a growing body of work that integrates multiple distinct clades within a given community (Webb *et al.* 2002; Cavender-Bares *et al.* 2009; Silvestro *et al.* 2015; Condamine *et al.* 2019). The model uses times of island colonisation and speciation events as well as the endemism status of the island species to estimate a set of parameters (cladogenesis rate  $\lambda^c$ , extinction rate  $\mu$ , colonisation rate  $\gamma$ , anagenesis rate  $\lambda^a$  and a carrying capacity  $K'$ ). The framework is founded on the birth-death model for reconstructed phylogenetic trees (i.e. only extant species) and can incorporate diversity-dependence in colonisation and speciation rates through a clade-specific (only within clades) or island-wide (across the entire island community) carrying capacity, as well as species known to exist but not present in the phylogeny (Etienne *et al.* 2012; 2022). DAISIE has been used for various studies in island biogeography. For example, it has allowed estimating rates of speciation, colonisation and extinction of the Galapagos terrestrial avifauna (Valente *et al.* 2015), it enabled the first global analysis of MacArthur & Wilson’s (1967) area and isolation model using phylogenetic data (Valente *et al.* 2020), it has put the impact of human-induced extinction in a macro-evolutionary context (Michielsen *et al.* 2023), and it has shown the presence of colonisation rate shifts of fishes in a Japanese Lake (Hauffe *et al.* 2020).

DAISIE allows studying diversification dynamics in insular communities without the need for a full phylogeny. Such a phylogeny is often not available, and the often employed alternative, focusing on the phylogeny of a single clade on the island, will not make use of the information from other lineages (small or large), and cannot be used to estimate colonisation rates. DAISIE uses a unique data structure: instead of considering a single insular lineage (e.g., the phylogeny of the Galapagos finches), it focuses on multiple independent lineages descending from different colonisation events of the island, representing various lineages that make up an insular community (e.g. phylogenies for each of the birds that colonised the Galapagos). Some of these lineages may have radiated into a large clade with many species on the island, others may have only a single species or few species on the island. Some of the species may be endemic to the island (found only on the island) and some may be non-endemic (also present outside of the island, i.e. the mainland). Phylogenetic data of sister species on the mainland are also needed to obtain an estimate of the colonisation time for each lineage. The data used by DAISIE thus includes colonisation and speciation times for the various lineages (obtained from time-calibrated molecular phylogenies), endemism status (endemic or non-endemic to the island) and the number of missing species (how many species from each insular lineage are not represented because molecular data is not available). A key feature of the model is that it does not require all the information on the different insular lineages to come from a single phylogeny. Instead, data from different phylogenetic trees can be used (for example, phylogenetic analyses of different genera or families from different publications).

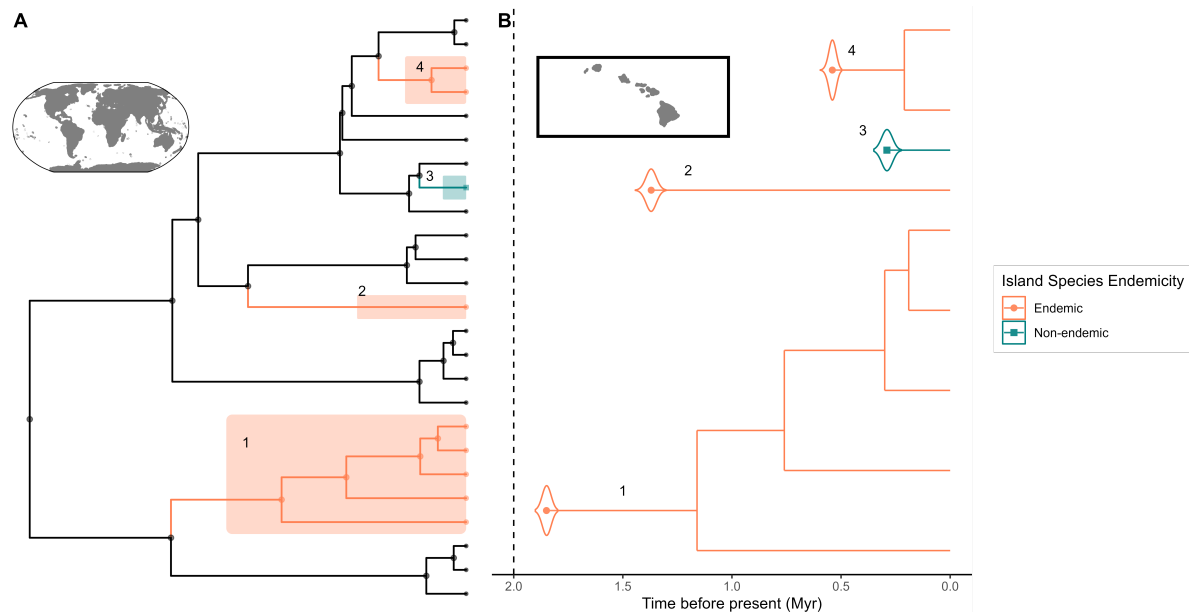


Figure 1: Visual representation of a typical dataset used in DAISIE analyses. In this hypothetical example, the focal insular system is the archipelago of Hawaii, and the focal island community has nine species. Panel A shows the wider global phylogeny in which the nine Hawaiian species are embedded (highlighted). Panel B shows the separate phylogenies of those island species from the perspective of the archipelago. Numbers on the plots link the corresponding lineages from each panel. The island community (B) consists of lineages resulting from four colonisation events, which are spread out in different topological locations in the tree of the taxonomic group at the global scale (A). Three of the island lineages contain exclusively species that are endemic to the island and one contains only a single non-endemic species. Two of the endemic lineages have undergone cladogenetic speciation on the island, forming island clades with more than one species. The colonisation time is assumed to be the divergence time from the mainland sister species (stem age). Violin plots around the island colonisation time show the uncertainty in divergence time estimates, representing a range of times from a posterior distribution of trees. The objective of the **DAISIEprep** package is to convert global phylogenetic data (A) to island community data (B).

The specific format of the data required to fit DAISIE models (Figure 1B) may constitute a barrier to its use. Here we present an R package, **DAISIEprep**, to provide a comprehensive set of tools to easily create data objects from phylogenetic data and island species checklists to be analysed in the maximum likelihood island biogeography model DAISIE (Lambert *et al.* 2023b). The package facilitates reproducible, fast extraction of community data from phylogenetic data (Figure 1). It is also possible to insert data on species or clades for which phylogenetic data is not available, but for which the number of species and endemicity statuses are known. If a species or island lineage is not sampled in a time-calibrated phylogeny, but an estimate for the time of island colonisation is available from other data sources, this can also be included. As software that produces phylogenetic data – for example BEAST2 (Bouckaert *et al.* 2014) and MrBayes (Huelsenbeck & Ronquist 2001) among others – output standard phylogenetic data structures, a package to convert these types of data into island-specific community data is a useful intermediate tool between primary empirical data and model fitting. Here, we explain the main functionalities of the R package across a range of scenarios, some auxiliary functions, as well as the downstream influence of data extraction choices.

## Package description

The package aims to facilitate and automate extracting and formatting data prior to applying DAISIE. This alleviates users having to write custom scripts. The **DAISIEprep** package requires as input a checklist of species on the island or archipelago of interest as well as time-calibrated phylogenetic data for these species. **DAISIEprep** uses phylogenetic data stored as a **phylo4d** object from the **phylobase** R package (Hackathon *et al.* 2020) which can handle phylogenetic trees with data at the tips and nodes of the tree. We also introduce three novel data structures (**Island\_colonist**, **Island\_tbl** and **Multi\_island\_tbl**) to aid in the handling of the specific community phylogenetic data, as well as a set of functions (methods) to easily modify the data, making use of the benefits of object-oriented programming in R (see Chambers (2014)). To check if the input data is compatible with **DAISIEprep**, the function **check\_phylo\_data()** can be used.

The package provides two algorithms to extract the community phylogenetic data: (1) minimum time of colonisation ('min'), and (2) ancestral state reconstruction ('asr'). The 'min' algorithm assumes that the stem node of an island lineage (the split from closest non-island or island non-endemic relatives sampled in the phylogeny) corresponds to the colonisation time of the island; and that there is no back-colonisation from the island to the mainland. The 'min' algorithm's assumption that species cannot migrate away from the island are consistent with the assumptions of the original DAISIE model. Under these assumptions, colonisation times are taken as the stem age. In the case when non-endemic island species or endemic island species with no sister species on the island are represented by a single tip in the phylogeny, the age of the node leading to the single tip corresponds to the colonisation time. When they are represented by several individuals or tips in the phylogeny, and in the case of endemic island clades with more than one species sampled in the phylogeny, the stem age of the island lineage corresponds to the colonisation time. This algorithm is useful when data seems to abide by the assumption of no back-colonisation or colonisation to other islands thus changing the endemism status to non-endemic. However, should such colonisation occur, the estimated number of island colonisations will be inflated, as island radiations may be artificially broken up into multiple colonisations of the island. In such cases, one may prefer to extract species based on a different model. To determine when colonisation happened, this would require a probabilistic mapping of island occupancy along the internal branches of the phylogeny, i.e. ancestral state reconstruction (see Joy *et al.* (2016) for review). The 'asr' algorithm is based on phylogenetic ancestral state reconstruction, and uses the ancestral geographical states at the nodes within the tree inferred with reconstruction methods to determine a likely number of colonisation events of the island. **DAISIEprep** provides the **add\_asr\_node\_states()** function to easily estimate ancestral states, using the **castor** R package to provide the maximum parsimony and the Mk model (Louca & Doebeli 2018).

The **DAISIEprep** package is set up to allow users to easily import ancestral range reconstructions obtained using methods from other packages, such as DEC/DEC+J (Matzke 2014; 2021), State-Speciation-Extinction models (Maddison *et al.* 2007; Goldberg *et al.* 2011; FitzJohn 2012; Herrera-Alsina *et al.* 2019), or stochastic character mapping (Huelsenbeck *et al.* 2003). More details are provided in the ancestral state reconstruction extensions vignette in the **DAISIEprep** package.

The 'asr' algorithm prevents clades being artificially split up in the case of a further colonisation event (Figure 2). In the extraction process the non-endemic is represented as an endemic species in the **Island\_tbl** and when fitting the DAISIE model. Therefore, the differences between the 'min' and 'asr' algorithms are both in the number of colonisation events inferred, as well as the endemism of species within island clades (Figure 2). Furthermore, under both 'min' and 'asr' algorithms, it is possible that the phylogenetic data or state reconstruction indicate a colonisation time older than the island age, for example if the closest relative of the island species has gone extinct or is not sampled in the phylogeny (Lambert *et al.* 2022). In such cases, the island lineage is assumed to have colonised any time after the island formation, with the DAISIE inference model integrating over this uncertainty.

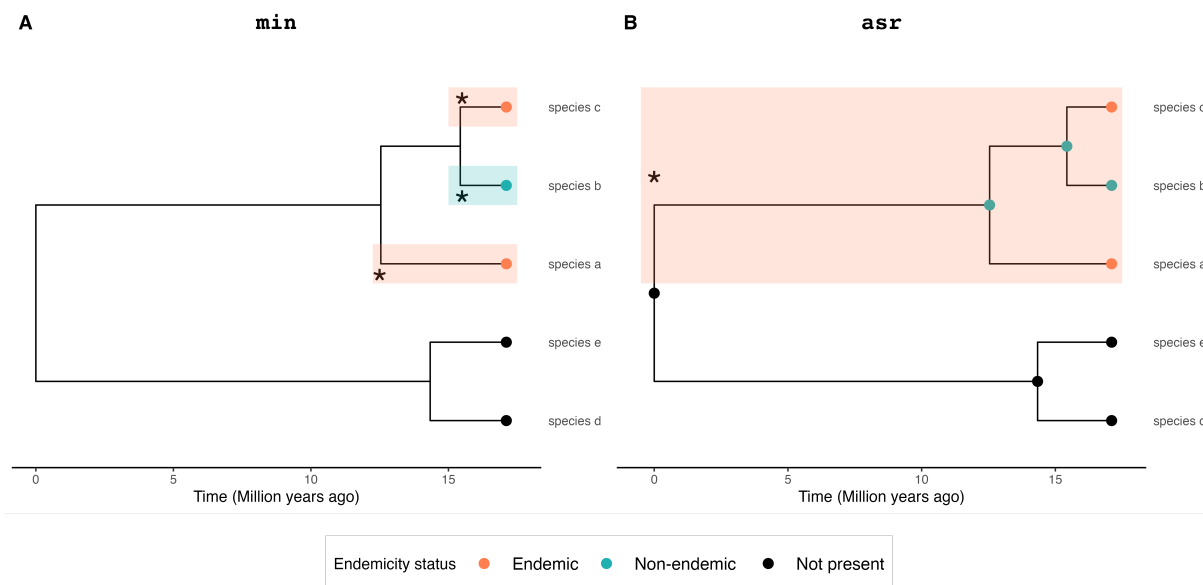


Figure 2: Comparison of the ‘min’ (A) and ‘asr’ (B) algorithms for data extraction. Two identical phylogenetic trees with five species, two of which are endemic, one non-endemic and two not present on the island. The ‘min’ algorithm (A) extracts three species, with stem times shown by \* and shaded areas, as there is a non-endemic species embedded within a clade with two endemic species. Thus using the ‘min’ algorithm assumes independent colonisations of each, minimising the time of colonisation. In contrast, the ‘asr’ algorithm (B) uses the internal node states (shown by dots at each node), to extract a single island colonist with three species, with the stem age of the island clade assumed to be the colonisation time (\* and shaded area). In this case (B), the extracted clade (shaded) is assumed endemic even though it contains a non-endemic species given the assumptions of the DAISIE inference model. Therefore the endemism status of the non-endemic is converted to endemic because the DAISIE inference model does not allow non-endemic species embedded within endemic island clades from a single colonisation because it does not consider colonisation from the island.

## Installation of DAISIEprep

The DAISIEprep package can be installed from github using the `remotes` R package (Csárdi *et al.* 2021) `remotes::install_github("joshwlambert/DAISIEprep")` and will be available on CRAN.

## Applications

An analysis using the DAISIE model starts with a checklist of the (native) species of the focal group that are found on an island, for example, all ferns or all squamates. This checklist should include whether each species is endemic to the island or not. The next step is to gather phylogenetic data from which colonisation and branching times of the island species can be inferred. The ‘ideal’ dataset would be a single time-calibrated phylogenetic tree with complete sampling, in which all species of the focal group are included (preferably multiple individuals of the same species) as well as all of their closest non-island relatives. In practice this type of dataset is very rare and difficult to produce. For most insular communities, there may be a few well-sampled phylogenies available for some of the insular lineages, while other lineages may have poorly sampled phylogenies, and for some lineages there may be no dated phylogenetic data available at all. For some non-endemic species there may be individuals from the mainland sampled in the phylogeny, but not from the island. DAISIE can incorporate information from these heterogeneous data types.

Here we demonstrate the basic usage of the DAISIEprep package through two applications to empirical data. The first is on a species-level macro-phylogeny of mammals by Upham *et al.* (2019), which we use to obtain information on the mammals of the island of Madagascar. This represents an application that may be common when users want to extract island data from a single global/macro-phylogeny containing a whole taxonomic group (e.g. birds (Jetz *et al.* 2012), or amphibians (Jetz & Pyron 2018)). The second empirical example is a data set of the Asteraceae plant family of the Hawaiian archipelago. Asteraceae form a hyper-diverse group of plants that is globally distributed on islands. Unlike for mammals, there is no species-level phylogenetic tree for the Asteraceae family (> 30,000 species). Thus, the Hawaiian Asteraceae dataset consists of multiple phylogenies for different Hawaiian genera or radiations, as well as species richness and endemism data for those Hawaiian clades for which no phylogenetic data are available (Landis *et al.* 2018; Knope *et al.* 2020; Keeley *et al.* 2021). Together, these two examples illustrate the use of macro-phylogenies, single or multiple phylogenies as well as the addition of island taxa when no phylogenetic data is available.

### Single phylogeny example: mammals of Madagascar

First, we used the checklist of all the mammal species that inhabit Madagascar from Michielsen *et al.* (2023), which recorded the island species' endemism status (endemic or not endemic to Madagascar). A table with this endemism information is a requirement for all uses of the DAISIEprep package. We link this table with the phylogenetic data using the `create_endemism_status()` function and the abovementioned `phylo4d` class (this creates the `phylod` object in the code example below). The mammal phylogeny used is the maximum clade credibility tree (Heled & Bouckaert 2013) from Upham *et al.* (2019). For this example we will use the 'asr' algorithm. The table (checklist of species with endemism statuses) and phylogeny are all the data required to extract the island community data using `extract_island_species()`.

```
# extract island community
island_tbl <- DAISIEprep::extract_island_species(
  phylod = phylod,
  extraction_method = "asr"
)
```

This outputs an `Island.tbl` object containing all the information on the colonisation and island speciation times, the endemism status, the number of missing species and name of each colonist. By default, the number of missing species for each island colonist is set to zero. To automatically count and assign missing species to the `Island.tbl` using information stored in the island species checklist, `count_missing_species()`, `unique_island_genera()` and `add_multi_missing_species()` can be used. Conversely, if missing species have been misassigned in this process, they can manually be removed from the `Island.tbl` using `rm_multi_missing_species()`. This set of functions assists in tabulating and assigning missing species to the island community data that was directly extracted from the phylogenetic tree (using `extract_island_species()`). To deal with the missing species remaining after automatic assignment, we provide a set of functions to add the missing species manually, in one of two ways, either assigning the missing species to an existing island clade (not to a specific topological location within the clade, but contributing to the total diversity of the clade) in the `Island.tbl` using `add_missing_species()`, or adding the missing species as a new separate island clade using `add_island_colonist()` (Figure 3). See the tutorial vignette in the DAISIEprep R package for worked examples of adding missing species under the different scenarios.



### Example of adding missing species to data

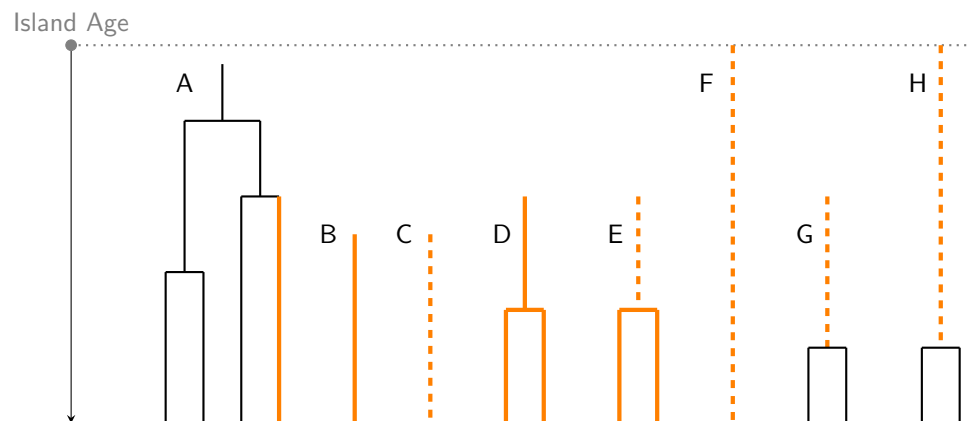


Figure 3: Adding missing species to island community data. The clade labelled A is a case when the phylogeny of an island clade is known (black lines), but one of the species is missing from the phylogeny (orange line). In this scenario the `add_missing_species()` function would be used to add, in this case, one missing species to the existing island clade in the island data set. The lineage labelled B is an island singleton (i.e. not an island radiation/clade) which is known to be on the island but has no time-calibrated phylogenetic data available. In this case an estimate of time of island colonisation exists, possibly from the literature. The lineage labelled C is the same as B except that the time taken from the literature is considered a maximum possible colonisation time rather than an exact time of colonisation (indicated by the dashed line). Island clade D is the same as B except that there is a clade instead of a singleton, in this case two species, on the island. Island clade E is the same as D but the colonisation time is a maximum and not an exact time (as in C). Lineage F is the case when a species is known on the island but its time of colonisation is unknown. Island clades G and H represent cases where the time of colonisation from the stem age is known or unknown, respectively, for a clade of two island species but an island speciation time is known (crown age) and can be used as a minimum time of colonisation. When the DAISIE model is fitted to the cases shown in C, E, F and H the model integrates between the maximum and minimum possible colonisation times (dashed lines) Examples B-H all use the `add_island_colonist()` to add island lineages the data set. The topology of missing species (island clade A) and their branching times (island clades D and E) are not known and are shown here just for illustrative purposes; see Etienne *et al.* (2012). Species endemism is not shown for simplicity. See the tutorial vignette in the R package for a full explanation of each case with examples.

When using `add_island_colonist()`, a colonisation can either be taken from the literature, extracted from the stem age of the genus if it is present in the phylogeny (i.e. the genus has mainland species in the tree) (Figure 3B, C, D, E), or can be given as unknown (using NA) (Figure 3F). An example of adding missing species in the case of the Malagasy mammals is the bat genus *Chaerephon*, which is missing and needs to be assigned, but it has species in the phylogeny not present on Madagascar. Thus a stem age can be extracted which is used as a maximum possible colonisation time. Setting `col_max_age = TRUE` specifies that the colonisation time given (in this case, the stem age) should be considered an upper bound rather than a precise time of colonisation.

```
# extract Chaerephon stem age from the phylogeny
Chaerephon_stem_age <- DAISIEprep::extract_stem_age(
  genus_name = "Chaerephon",
  phylod = phylod,
  stem = "genus"
)
```

```
# add the Chaerephon with a maximum colonisation time as its stem age
island_tbl <- DAISIEprep::add_island_colonist(
  island_tbl = island_tbl,
  clade_name = "Chaerephon_leucogaster",
  status = "nonendemic",
  missing_species = 0,
  col_time = Chaerephon_stem_age,
  col_max_age = TRUE,
  branching_times = NA_real_,
  min_age = NA_real_,
  species = "Chaerephon_leucogaster",
  clade_type = 1
)
```

Once the `Island.tbl` object is finalised it can be converted to the DAISIE data list structure required by the DAISIE likelihood model, using `create_daisie_data()` function. The script and phylogenetic data used for the mammals of Madagascar in this section for loading the checklist, extracting the data, and handling the missing species can be found in the `DAISIEprepExtra` R package (Lambert *et al.* 2023a).

### Multiple phylogenies example: Hawaiian Asteraceae

The case of the Asteraceae species of the archipelago of Hawaii provides a good example of the spectrum of phylogenetic and diversity data available for an island community. The Hawaiian archipelago currently has 98 native species (97 endemic) of Asteraceae, which are the result of at least 10 colonization events. There are several new species that are currently awaiting description, so this number will change in the future, but for the purposes of this example we use the currently accepted species. Each of the Hawaiian Asteraceae clades has received varying degrees of attention, contributing to differing data availability across the community. For example, the Hawaiian Silverswords and Pacific *Bidens* – two classic examples of adaptive radiations – are well-studied with dated phylogenies available; whereas, *Tetramolopium* – a potential radiation of 11 species – does not have a phylogeny available.

We compiled a checklist of the Asteraceae species native and endemic to Hawaii and gathered all available phylogenetic data. For the 10 assumed colonization events, three clades (representing 60.5% of the species) have dated phylogenies publicly available (Hawaiian Silversword alliance, Landis *et al.* (2018); *Bidens*, Knope *et al.* (2020); *Hesperomannia*, Keeley *et al.* (2021)); three clades have dated phylogenies but tree files are not publicly available, so we use the colonisation time estimates cited in the text of the publications (*Artemisia*, Hobbs & Baldwin (2013); *Lipochaeta-Melanthera* alliance, Edwards *et al.* (2018); *Pseudognaphalium*, Nie *et al.* (2016)); and four clades do not have a published phylogeny, and we use the diversity data (number of species thought to belong to the same insular lineage based on taxonomic information, and endemism status) (*Tetramolopium*, *Keysseria*, *Remya*, *Adenostemma*). The resulting community dataset leverages multiple dated phylogenies, colonisation time estimates provided in-text in scientific articles, and island diversity data (number of species and endemism status) for cases for which no phylogenetic work has been done on the island species or clades (Figure 4). For the cases where only crown ages were available, these were used as the minimum colonisation times (most recent possible time of colonisation).

The pipeline to extract data from multiple phylogenies is the same as for the single phylogeny example, but now `extract_island_species()` is called for each of the three available phylogenies (i.e. Hawaiian Silversword alliance, *Bidens*, and *Hesperomannia*), with the argument `island.tbl` to bind to the same `island.tbl` object.

```
# add silverswords to existing island.tbl
```



```
island_tbl <- DAISIEprep::extract_island_species(
  phylod = silversword_phylod,
  extraction_method = "min",
  island_tbl = island_tbl
)
```

Island colonists that do not have phylogenetic data, but may have a known stem and crown age, are input using the function `add_island_colonist()`. Here the genus *Artemisia* is inserted with a stem and crown age estimate obtained from the literature (Hobbs & Baldwin 2013). The *Lipochaeta-Melanthera* alliance are inserted with a stem age estimated from the literature (Edwards *et al.* 2018) but unknown crown age.

```
# add Artemisia to existing island_tbl; we have the stem age (3.93 Ma) and crown
# (1.45 Ma) from Hobbs and Baldwin (2013)
island_tbl <- DAISIEprep::add_island_colonist(
  island_tbl = island_tbl,
  clade_name = "Artemisia",
  status = "endemic",
  missing_species = 1,
  col_time = 3.93,
  col_max_age = TRUE,
  branching_times = 1.45,
  min_age = NA_real_,
  species = c(
    "Artemisia_kauaiensis",
    "Artemisia_mauiensis",
    "Artemisia_australis"
  ),
  clade_type = 1
)
```

Lastly, when no phylogenetic data is available for a taxonomic group and there exists no estimate for the colonisation time, a clade can be added with the maximum time of colonisation as the age of the island, and the number of species inserted from diversity data (Figure 4). This is done for the *Keysseria* clade in our example.

```
# add Keysseria
# diversity data from taxonomic sources
island_tbl <- DAISIEprep::add_island_colonist(
  island_tbl = island_tbl,
  clade_name = "Keysseria",
  status = "endemic",
  missing_species = 2,
  col_time = NA_real_,
  col_max_age = TRUE,
  branching_times = NA_real_,
  min_age = NA_real_,
  species = c(
    "Keysseria_maviensis",
    "Keysseria_erici",
    "Keysseria_helena"
  ),
  clade_type = 1
)
```

Once a final `Island_tbl` is compiled, `create_daisie_data()` can be used to convert the data for application in DAISIE.

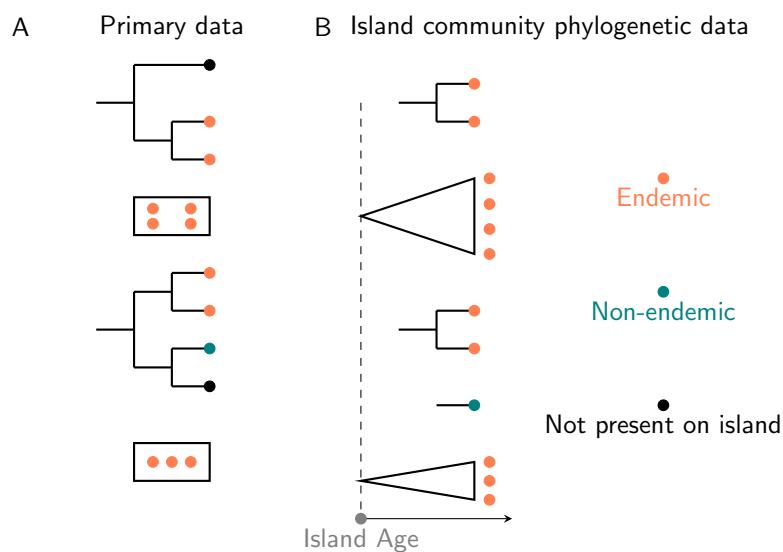


Figure 4: Example of using different sources and formats of data to obtain island community phylogeny data. In this case, the primary data (A) is composed of two phylogenetic trees that include some of the island species and, for cases where phylogenetic data is not available (represented by rectangular boxes), diversity data (number of species present on the island assumed to belong to the same insular lineage, and their endemism status). This example shows how island species or clades that are missing phylogenetic data but for which the island species diversity and endemism are known (shown in boxes), can be accounted for in the DAISIE framework (B) by considering them as missing species and treating the clade as a polytomy with known species richness and species endemism (shown by the triangles with the dots representing how many species are in those clades). The dashed line represents the age of the island, and those island species that lack phylogenetic data are assumed to have colonised the island any time after island formation.

## Posterior distribution example

Phylogenies are estimates of evolutionary history with uncertainty in both branching times and topology, which, in Bayesian analyses, can be represented in a posterior distribution (Huelsenbeck *et al.* 2000; 2001). When available, posterior distributions are often used to repeat downstream analyses on multiple sampled phylogenies, and account for the effect of phylogenetic uncertainty on the conclusions. DAISIEprep offers tools to explore the effect of phylogenetic uncertainty on the inferred colonisation and diversification history of the island, by handling the posterior distribution of phylogenetic trees. Here we use the posterior (or ‘pseudo-posterior’, Kuhn *et al.* (2011)) distribution of mammal phylogenies from Upham *et al.* (2019). The island data is individually processed for each phylogeny from the posterior, using `multi_extract_island_species()`, which returns a `Multi_island_tbl` object.

```
# extract island community using asr algorithm
multi_island_tbl <- DAISIEprep::multi_extract_island_species(
  multi_phylod = multi_phylod,
  extraction_method = "asr",
  verbose = TRUE
)
```

Thus multiple island data sets that reflect phylogenetic uncertainty are produced, with each providing an evolutionary hypothesis for the colonisation and diversification history of the island. Missing species are handled in the same way as for the single phylogeny example that used the maximum clade credibility mammal phylogeny. Finally, `create_daisie_data()` and `DAISIE::DAISIE_ML_CS()` can be looped over each island data set in order to incorporate the phylogenetic uncertainty. The full script is supplied in the `DAISIEprepExtra` R package.

## Utility functions

The package contains a set of utility functions to inspect the data and check for certain characteristics that may influence the choice of extraction algorithm ('min' or 'asr'). The plotting functions `plot_phylo()` and `plot_colonisation()` are for plotting the data before and after extraction, respectively using the `phylo` object (see above) containing phylogenetic and endemism data. For inspecting the phylogenetic tree and its associated data, either at the tips, or both at the tips and the internal nodes (Figure 5A), `plot_phylo` provides a way to visualise the distribution of island endemism across each tree and infer when island colonisation likely occurred. `plot_colonisation` displays for each island colonist lineage their stem age, and if available, the crown age (Figure 5B). `DAISIEprep` uses `ggplot2` (Wickham 2016) and `ggtree` (Yu *et al.* 2017; 2018) to produce plots of the island presence/absence at the tips of the phylogeny, the ancestral states at the nodes and other useful visualisation to aid the user in understanding the data extraction. By building on the flexibility of the `ggplot2` framework `DAISIEprep` plots can be modified using `ggplot2` layers.

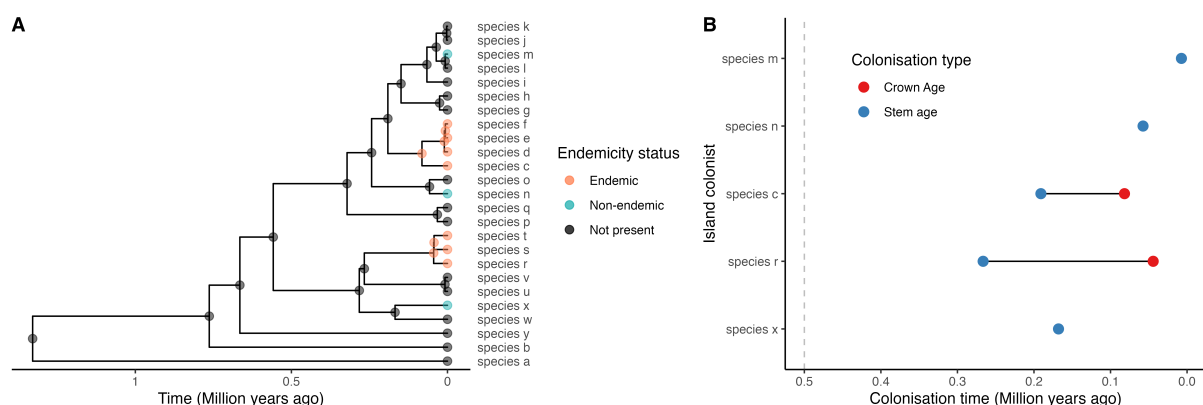


Figure 5: Demonstration of plots produced using (A) `plot_phylo()` and (B) `plot_colonisation()` for the same data. The example shows five island colonists, two of which are island radiations and thus have a crown age as well as a stem age shown in panel B. The dashed line in panel B is the island age.

Another utility function is the ability to check for back-or-onward-colonisation (`any_back_colonisation()`) events within the phylogeny given an ancestral state reconstruction of the island presence. Identifying such colonisations of island species back to the mainland or to another island (thereby changing the endemism status to non-endemic) can be useful to understand the dispersal patterns of a group and may provide insight into the prevalence of so-called boomerang colonisations (Bellemain & Ricklefs 2008). Identifying these colonisations can be challenging given the inherent limitation of reconstructing species ranges over millions of years, and the specified ancestral state reconstruction model and transition matrix may erroneously produce or miss such colonisations.

Finally, a check for any polyphyly (`any_polyphyly()`) on the tree is implemented in the package, to identify cases in which multiple samples of conspecific species are not found to be monophyletic in the

tree. This could for example result from low phylogenetic resolution or incomplete lineage sorting of recently colonised island species, or may indicate true polyphyly (e.g. multiple colonisation events of the island by the same species). When polyphyly of a species is identified, the current extraction algorithms cannot correctly extract these data, so the user needs to decide how to assign colonisation events for that species beforehand.

## Sensitivity of parameter estimates to data extraction method

The objective of using DAISIEprep as a processing tool in the pipeline of island biogeography analyses is most often going to be to estimate parameters of colonisation, speciation and extinction dynamics on islands. However, given the different extraction algorithms ('min' and 'asr') as well as the options within the 'asr' algorithm (e.g., model of ancestral range reconstruction and possible geographical transition matrix specified), there are many combinations of extraction settings. These may in some cases produce identical island data sets, while in others they may differ in the number of island colonisation events or branching times. We performed a simple sensitivity analysis (*sensu* global sensitivity, Saltelli *et al.* (2008)) to highlight how data extraction choice can have downstream effects in the parameter estimates obtained by DAISIE. We recommend users to carry out sensitivity analysis tailored to their data, in order to understand how the data extraction method influences outcomes on parameter estimation and model selection. Sensitivity to uncertainty in the island age and the number of mainland species should also be tested. See 'Sensitivity' vignette in the DAISIEprep package for more details and instructions for running sensitivity analyses.

To illustrate the impact of extraction methods on parameter estimates, we used 100 phylogenies from the posterior of the macro-phylogeny of mammals used in the first example above, focusing again on the species of Madagascar. We extracted the data using the 'min' and 'asr' methods. For the 'asr' method we use two approaches, maximum parsimony and Mk model (assuming equal transition rates) using the R package *castor*. We then applied a diversity-dependent DAISIE model to the resulting datasets, optimising all five parameters. Figure 6 shows that the estimates of anagenesis, cladogenesis and extinction largely overlap, but colonisation rate differs between the methods. This is because the 'asr' algorithm produces fewer colonisation events.

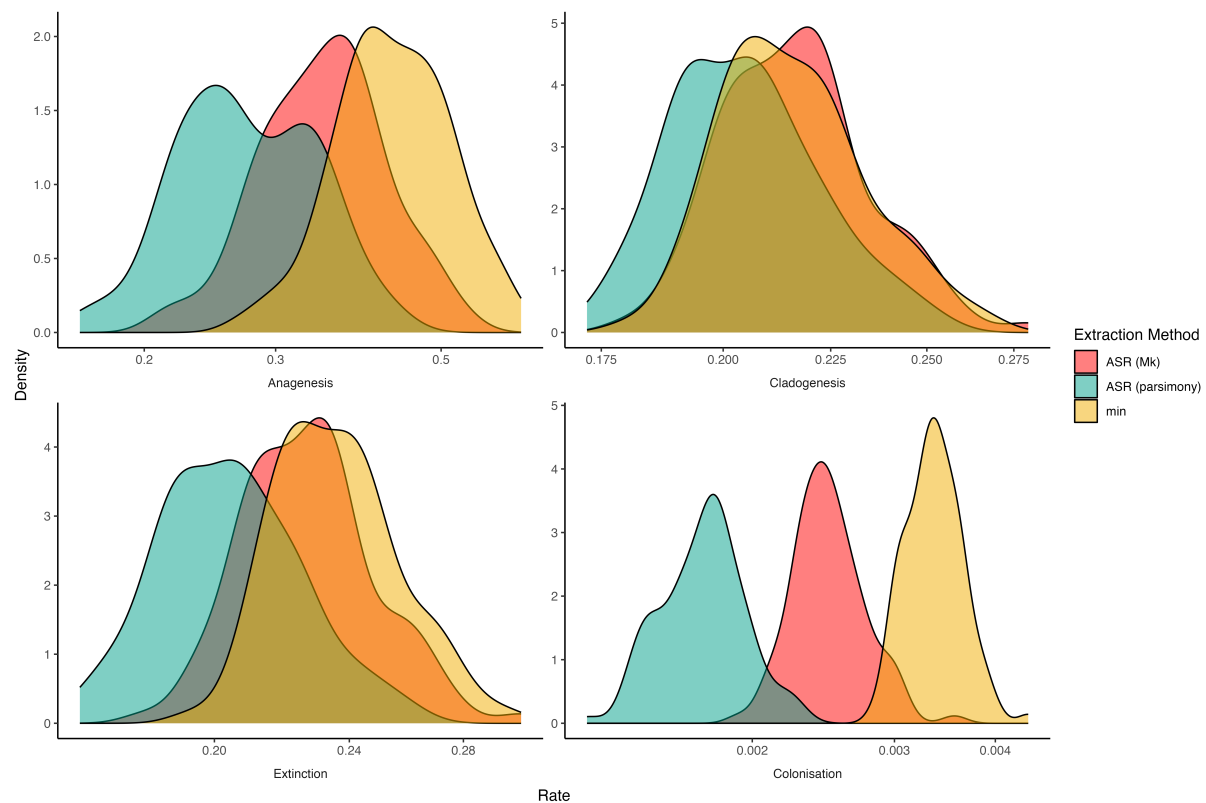


Figure 6: Parameter estimates from the DAISIE maximum likelihood model when data is extracted using different algorithms ('asr' and 'min') and variants of the 'asr' algorithm (parsimony and Mk model). Based on fitting the DAISIE model to 100 datasets (from the posterior distribution of the mammal macro-phylogeny) extracted under each of the extraction algorithms.

## Discussion

Previously, the process of compiling phylogenetic island data to fit community island biogeography models required manually locating island clades. Here we have presented the **DAISIEprep** R package, which automates this procedure and thus allows for a more seamless pipeline to go from commonly used phylogenetic data to model fitting. The examples presented above demonstrate the versatility of the package with a variety of input data that can be handled. The package can be used both on newly inferred phylogenies from sequence data, or previously published phylogenies, for example from a database of phylogenies (e.g. TreeBase, Piel *et al.* (2000)) or macro-phylogenies (Jetz *et al.* 2012; Jetz & Pyron 2018; Upham *et al.* 2019). The performance of the central function (`extract_island_species()`), in terms of computational time is found to be reasonable for common empirical trees (10 - 10,000 tips) with extraction taking seconds. The time consumption of computing the ancestral reconstruction or estimating DAISIE maximum likelihood parameters is on the same time scale or greater, and so this pre-processing tool does not constitute a bottleneck to the pipeline.

Even with this wealth of phylogenetic data there are still many cases for which no phylogenetic data is available or cannot be accessed (Magee *et al.* 2014). In such cases, the absence of a phylogeny can be mitigated by **DAISIEprep**, using island diversity and endemism data in a flexible framework. Using as much phylogenetic data available as possible is evidently still preferred, as it does improve inference reliability (Valente *et al.* 2018), and it improves knowledge on the number of colonisation events.

Both algorithms introduced in **DAISIEprep** ('min' and 'asr') include ancestral state reconstruction, either implicitly or explicitly. **DAISIEprep** has been set up to allow for alternative methods of ancestral state reconstruction methods. Therefore, if it is perceived that species range, either on the island, on the mainland or both, influences the rates of speciation and extinction, then a model that can account for these causes of rate heterogeneity can be used (Maddison 2006; Maddison *et al.* 2007; Holland *et al.* 2020). However, it is worth noting the known limitations of these methods. Particularly, they are highly uncertain further back in time (Cunningham 1999; Martins 2000; O'Meara & Beaulieu 2021). One encouraging aspect is that islands – especially oceanic islands – are usually young (5-10 Mya) and so species will have colonised relatively recently, making the inference of species arrival more accurate than if it were deeper in the past. Another limitation is the use of phylogenies containing species for which no molecular data is available and which are added to the trees using taxonomic constraints. Such methods have been shown to yield reliable, albeit conservative, conclusions from diversification analyses (Chang *et al.* 2020), yet, these methods may scramble the phylogenetic signal of island presence, especially when these missing species are only constrained at higher taxonomic levels of family or order (Rabosky 2015). In turn, this may potentially lead to an increase in island colonisation events as species from the same insular radiation are incorrectly placed in different locations across the tree. In these cases, results should be validated with the equivalent DNA-only phylogeny.

By releasing this open-source package we hope to encourage the use, development and scrutiny of phylogenetic inference models, and allow the community of phylogenetic inference to expand and answer more questions in the fields of biogeography and community assembly, while providing high reproducibility. This package will continue development in parallel with the **DAISIE** R package.

## Acknowledgments

We would like to thank the Theoretical and Evolutionary Community Ecology group at the University of Groningen, particularly Pedro Santos Neves for discussions and their input. We would like to thank the Center for Information Technology of the University of Groningen for their support and for providing access to the Peregrine high performance computing cluster. JWJ was funded through a Study Abroad



Studentship by the Leverhulme Trust and was also funded by a NWO VICI grant awarded to RSE who was funded by the same grant. TP was funded by the University of Groningen and the Faculty of Natural Sciences, University of Stirling (UK). LR was funded by an NWO VIDI grant awarded to LV, LV was funded by the same grant. All authors declare no conflict of interest.

## Authors contributions

JWL formulated the project, drafted the manuscript, created and led development on the DAISIEprep package, and ran the analyses. LR collected the Hawaiian Asteraceae data and assisted in writing the R script and manuscript on the multiple phylogeny example. TP wrote the extending ancestral state reconstruction vignette in the R package. LV collected the data for the Malagasy Mammals (originally from Michielsen *et al.* (2023)) and assisted in the development of the R package, as well as supervised the project. RSE supervised the project and gave feedback throughout. All authors revised the manuscript and approved publication.

## References

- Bellemain, E. & Ricklefs, R. (2008) Are islands the end of the colonization road? *Trends in Ecology & Evolution*, **23**, 461–468. <https://dx.doi.org/10.1016/j.tree.2008.05.001>.
- Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.H., Xie, D., Suchard, M.A., Rambaut, A. & Drummond, A.J. (2014) BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. *PLoS Computational Biology*, **10**, e1003537. <https://dx.doi.org/10.1371/journal.pcbi.1003537>.
- Cardona, G., Rosselló, F. & Valiente, G. (2008) Extended Newick: it is time for a standard representation of phylogenetic networks. *BMC Bioinformatics*, **9**, 532. <https://dx.doi.org/10.1186/1471-2105-9-532>.
- Cavender-Bares, J., Kozak, K.H., Fine, P.V.A. & Kembel, S.W. (2009) The merging of community ecology and phylogenetic biology. *Ecology Letters*, **12**, 693–715. <https://dx.doi.org/10.1111/j.1461-0248.2009.01314.x>.
- Chambers, J.M. (2014) Object-Oriented Programming, Functional Programming and R. *Statistical Science*, **29**. <https://dx.doi.org/10.1214/13-STS452>.
- Chang, J., Rabosky, D.L. & Alfaro, M.E. (2020) Estimating Diversification Rates on Incompletely Sampled Phylogenies: Theoretical Concerns and Practical Solutions. *Systematic Biology*, **69**, 602–611. <https://dx.doi.org/10.1093/sysbio/syz081>.
- Condamine, F.L., Romieu, J. & Guinot, G. (2019) Climate cooling and clade competition likely drove the decline of lamniform sharks. *Proceedings of the National Academy of Sciences*, **116**, 20584–20590. <https://dx.doi.org/10.1073/pnas.1902693116>.
- Csárdi, G., Hester, J., Wickham, H., Chang, W., Morgan, M. & Tenenbaum, D. (2021) remotes: R Package Installation from Remote Repositories, Including 'GitHub'.
- Cunningham, C.W. (1999) Some Limitations of Ancestral Character State Reconstruction When Testing Evolutionary Hypotheses. *Systematic Biology*, **48**, 665–674.
- Edwards, R.D., Cantley, J.T., Chau, M.M., Keeley, S.C. & Funk, V.A. (2018) Biogeography and relationships within the *Melanthera* alliance: A pan-tropical lineage (Compositae: Heliantheae: Ecliptinae). *Taxon*, **67**, 552–564. <https://dx.doi.org/10.12705/673.6>.
- Etienne, R.S., Haegeman, B., Dugo-Cota, A., Vilà, C., Gonzalez-Voyer, A. & Valente, L. (2022) The limits to ecological limits to diversification. preprint, Evolutionary Biology.

- Etienne, R.S., Haegeman, B., Stadler, T., Aze, T., Pearson, P.N., Purvis, A. & Phillimore, A.B. (2012) Diversity-dependence brings molecular phylogenies closer to agreement with the fossil record. *Proceedings of the Royal Society B: Biological Sciences*, **279**, 1300–1309. <https://dx.doi.org/10.1098/rspb.2011.1439>.
- Felsenstein, J. (2004) *Inferring phylogenies*. Sinauer Associates, Sunderland, Mass.
- FitzJohn, R.G. (2012) Diversitree : comparative phylogenetic analyses of diversification in R: *Diversitree*. *Methods in Ecology and Evolution*, **3**, 1084–1092. <https://dx.doi.org/10.1111/j.2041-210X.2012.00234.x>.
- Goldberg, E.E., Lancaster, L.T. & Ree, R.H. (2011) Phylogenetic Inference of Reciprocal Effects between Geographic Range Evolution and Diversification. *Systematic Biology*, **60**, 451–465. <https://dx.doi.org/10.1093/sysbio/syr046>.
- Hackathon, R., Bolker, B., Butler, M., Cowan, P., de Vienne, D., Eddelbuettel, D., Holder, M., Jombart, T., Kembel, S., Michonneau, F., Orme, D., O'Meara, B., Paradis, E., Regetz, J. & Zwickl, D. (2020) phylobase: Base Package for Phylogenetic Structures and Comparative Data.
- Hauffe, T., Delicado, D., Etienne, R.S. & Valente, L. (2020) Lake expansion elevates equilibrium diversity via increasing colonization. *Journal of Biogeography*, **47**, 1849–1860. <https://dx.doi.org/10.1111/jbi.13914>.
- Heled, J. & Bouckaert, R.R. (2013) Looking for trees in the forest: summary tree from posterior samples. *BMC Evolutionary Biology*, **13**, 221. <https://dx.doi.org/10.1186/1471-2148-13-221>.
- Herrera-Alsina, L., van Els, P. & Etienne, R.S. (2019) Detecting the Dependence of Diversification on Multiple Traits from Phylogenetic Trees and Trait Data. *Systematic Biology*, **68**, 317–328. <https://dx.doi.org/10.1093/sysbio/syy057>.
- Hobbs, C.R. & Baldwin, B.G. (2013) Asian origin and upslope migration of Hawaiian *Artemisia* (Compositae-Anthemideae). *Journal of Biogeography*, **40**, 442–454. <https://dx.doi.org/10.1111/jbi.12046>.
- Holland, B.R., Ketelaar-Jones, S., O'Mara, A.R., Woodhams, M.D. & Jordan, G.J. (2020) Accuracy of ancestral state reconstruction for non-neutral traits. *Scientific Reports*, **10**, 7644. <https://dx.doi.org/10.1038/s41598-020-64647-4>.
- Huelsenbeck, J.P. & Ronquist, F. (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*, **17**, 754–755. <https://dx.doi.org/10.1093/bioinformatics/17.8.754>.
- Huelsenbeck, J.P., Nielsen, R. & Bollback, J.P. (2003) Stochastic Mapping of Morphological Characters. *Systematic Biology*, **52**, 131–158. <https://dx.doi.org/10.1080/10635150390192780>.
- Huelsenbeck, J.P., Rannala, B. & Masly, J.P. (2000) Accommodating Phylogenetic Uncertainty in Evolutionary Studies. *Science*, **288**, 2349–2350. <https://dx.doi.org/10.1126/science.288.5475.2349>.
- Huelsenbeck, J.P., Ronquist, F., Nielsen, R. & Bollback, J.P. (2001) Bayesian Inference of Phylogeny and Its Impact on Evolutionary Biology. *Science*, **294**, 2310–2314. <https://dx.doi.org/10.1126/science.1065889>.
- Jetz, W., Thomas, G.H., Joy, J.B., Hartmann, K. & Mooers, A.O. (2012) The global diversity of birds in space and time. *Nature*, **491**, 444–448. <https://dx.doi.org/10.1038/nature11631>.
- Jetz, W. & Pyron, R.A. (2018) The interplay of past diversification and evolutionary isolation with present imperilment across the amphibian tree of life. *Nature Ecology & Evolution*, **2**, 850–858. <https://dx.doi.org/10.1038/s41559-018-0515-5>.

- Joy, J.B., Liang, R.H., McCloskey, R.M., Nguyen, T. & Poon, A.F.Y. (2016) Ancestral Reconstruction. *PLOS Computational Biology*, **12**, e1004763. <https://dx.doi.org/10.1371/journal.pcbi.1004763>.
- Keeley, S.C., Cantley, J.T. & Gallaher, T.J. (2021) The “evil tribe” spreads across the land: A dated molecular phylogeny provides insight into dispersal, expansion, and biogeographic relationships within one of the largest tribes of the sunflower family (Vernonieae: Compositae). *American Journal of Botany*, **108**, 505–519. <https://dx.doi.org/10.1002/ajb2.1614>.
- Knape, M.L., Bellinger, M.R., Datlof, E.M., Gallaher, T.J. & Johnson, M.A. (2020) Insights into the Evolutionary History of the Hawaiian *Bidens* (Asteraceae) Adaptive Radiation Revealed Through Phylogenomics. *Journal of Heredity*, **111**, 119–137. <https://dx.doi.org/10.1093/jhered/esz066>.
- Kuhn, T.S., Mooers, A.O. & Thomas, G.H. (2011) A simple polytomy resolver for dated phylogenies. *Methods in Ecology and Evolution*, **2**, 427–436. <https://dx.doi.org/10.1111/j.2041-210X.2011.00103.x>.
- Lambert, J., Valente, L. & Roebler, L. (2023a) DAISIEprepExtra: Scripts and documents to accompany and reproduce Lambert et al.
- Lambert, J.W., Neves, P.S., Bilderbeek, R.J.C., Valente, L. & Etienne, R.S. (2022) The effect of mainland dynamics on data and parameter estimates in island biogeography. preprint, Evolutionary Biology.
- Lambert, J.W., Valente, L., Roebler, L., Pannetier, T. & Neves, P.S. (2023b) DAISIEprep: Extracts phylogenetic island community data from phylogenetic trees. Language: en.
- Landis, M.J., Freyman, W.A. & Baldwin, B.G. (2018) Retracing the Hawaiian silversword radiation despite phylogenetic, biogeographic, and paleogeographic uncertainty. *Evolution*, **72**, 2343–2359. <https://dx.doi.org/10.1111/evo.13594>.
- Louca, S. & Doebeli, M. (2018) Efficient comparative phylogenetics on large trees. *Bioinformatics*, **34**, 1053–1055. <https://dx.doi.org/10.1093/bioinformatics/btx701>.
- MacArthur, R.H. & Wilson, E.O. (1967) *The theory of island biogeography*. Princeton University Press, Princeton.
- Maddison, D.R., Swofford, D.L. & Maddison, W.P. (1997) Nexus: An Extensible File Format for Systematic Information. *Systematic Biology*, **46**, 590–621. <https://dx.doi.org/10.1093/sysbio/46.4.590>.
- Maddison, W.P. (2006) Confounding asymmetries in evolutionary diversification and character change. *Evolution*, **60**, 1743–1746. <https://dx.doi.org/10.1111/j.0014-3820.2006.tb00517.x>.
- Maddison, W.P., Midford, P.E. & Otto, S.P. (2007) Estimating a Binary Character’s Effect on Speciation and Extinction. *Systematic Biology*, **56**, 701–710. <https://dx.doi.org/10.1080/10635150701607033>.
- Magee, A.F., May, M.R. & Moore, B.R. (2014) The Dawn of Open Access to Phylogenetic Data. *PLoS ONE*, **9**, e110268. <https://dx.doi.org/10.1371/journal.pone.0110268>.
- Martins, E.P. (2000) Adaptation and the comparative method. *Trends in Ecology & Evolution*, **15**, 296–299. [https://dx.doi.org/10.1016/S0169-5347\(00\)01880-2](https://dx.doi.org/10.1016/S0169-5347(00)01880-2).
- Matzke, N. (2021) Statistical comparison of DEC and DEC+J is identical to comparison of two ClaSSE submodels, and is therefore valid. preprint, Open Science Framework.
- Matzke, N.J. (2014) Model Selection in Historical Biogeography Reveals that Founder-Event Speciation Is a Crucial Process in Island Clades. *Systematic Biology*, **63**, 951–970. <https://dx.doi.org/10.1093/sysbio/syu056>.

- Michielsen, N.M., Goodman, S.M., Soarimalala, V., van der Geer, A.A.E., Dávalos, L.M., Saville, G.I., Upham, N. & Valente, L. (2023) The macroevolutionary impact of recent and imminent mammal extinctions on Madagascar. *Nature Communications*, **14**, 14. <https://dx.doi.org/10.1038/s41467-022-35215-3>.
- Morlon, H., Lewitus, E., Condamine, F.L., Manceau, M., Clavel, J. & Drury, J. (2016) RPANDA: an R package for macroevolutionary analyses on phylogenetic trees. *Methods in Ecology and Evolution*, **7**, 589–597. <https://dx.doi.org/10.1111/2041-210X.12526>.
- Nie, Z., Funk, V.A., Meng, Y., Deng, T., Sun, H. & Wen, J. (2016) Recent assembly of the global herbaceous flora: evidence from the paper daisies (Asteraceae: Gnaphalieae). *New Phytologist*, **209**, 1795–1806. <https://dx.doi.org/10.1111/nph.13740>.
- O’Meara, B. & Beaulieu, J.M. (2021) Potential survival of some, but not all, diversification methods. preprint, EcoEvoRxiv.
- Paradis, E., Claude, J. & Strimmer, K. (2004) APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics*, **20**, 289–290. <https://dx.doi.org/10.1093/bioinformatics/btg412>.
- Pennell, M.W., Eastman, J.M., Slater, G.J., Brown, J.W., Uyeda, J.C., FitzJohn, R.G., Alfaro, M.E. & Harmon, L.J. (2014) geiger v2.0: an expanded suite of methods for fitting macroevolutionary models to phylogenetic trees. *Bioinformatics*, **30**, 2216–2218. <https://dx.doi.org/10.1093/bioinformatics/btu181>.
- Piel, H., W., Donoghue, M.J. & Sanderson, M.J. (2000) TreeBASE: a database of phylogenetic information. *Proceedings of the 2nd International Workshop of Species*.
- R Core Team (2022) R: A Language and Environment for Statistical Computing.
- Rabosky, D.L. (2015) No substitute for real data: A cautionary note on the use of phylogenies from birth-death polytomy resolvers for downstream comparative analyses. *Evolution*, **69**, 3207–3216. <https://dx.doi.org/10.1111/evo.12817>.
- Revell, L.J. (2012) phytools: an R package for phylogenetic comparative biology (and other things): *phytools: R package*. *Methods in Ecology and Evolution*, **3**, 217–223. <https://dx.doi.org/10.1111/j.2041-210X.2011.00169.x>.
- Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M. & Tarantola, S. (2008) *Global sensitivity analysis: the primer*. Wiley, New York. OCLC: 1073614520.
- Silvestro, D., Antonelli, A., Salamin, N. & Quental, T.B. (2015) The role of clade competition in the diversification of North American canids. *Proceedings of the National Academy of Sciences*, **112**, 8684–8689. <https://dx.doi.org/10.1073/pnas.1502803112>.
- Upham, N.S., Esselstyn, J.A. & Jetz, W. (2019) Inferring the mammal tree: Species-level sets of phylogenies for questions in ecology, evolution, and conservation. *PLOS Biology*, **17**, e3000494. <https://dx.doi.org/10.1371/journal.pbio.3000494>.
- Valente, L., Phillimore, A.B. & Etienne, R.S. (2015) Equilibrium and non-equilibrium dynamics simultaneously operate in the Galápagos islands. *Ecology Letters*, **18**, 844–852. <https://dx.doi.org/10.1111/ele.12461>.
- Valente, L., Phillimore, A.B. & Etienne, R.S. (2018) Using molecular phylogenies in island biogeography: it’s about time. *Ecography*, **41**, 1684–1686. <https://dx.doi.org/10.1111/ecog.03503>.
- Valente, L., Phillimore, A.B., Melo, M., Warren, B.H., Clegg, S.M., Havenstein, K., Tiedemann, R., Illera, J.C., Thébaud, C., Aschenbach, T. & Etienne, R.S. (2020) A simple dynamic model explains the diversity of island birds worldwide. *Nature*, **579**, 92–96. <https://dx.doi.org/10.1038/s41586-020-2022-5>.

- Webb, C.O., Ackerly, D.D., McPeck, M.A. & Donoghue, M.J. (2002) Phylogenies and Community Ecology. *Annual Review of Ecology and Systematics*, **33**, 475–505. <https://dx.doi.org/10.1146/annurev.ecolsys.33.010802.150448>.
- Wickham, H. (2016) *ggplot2: elegant graphics for data analysis*. Use R! Springer, Switzerland, second edition.
- Yu, G., Lam, T.T.Y., Zhu, H. & Guan, Y. (2018) Two Methods for Mapping and Visualizing Associated Data on Phylogeny Using *ggtree*. *Molecular Biology and Evolution*, **35**, 3041–3043. <https://dx.doi.org/10.1093/molbev/msy194>.
- Yu, G., Smith, D.K., Zhu, H., Guan, Y. & Lam, T.T. (2017) ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution*, **8**, 28–36. <https://dx.doi.org/10.1111/2041-210X.12628>.