Appendix A

Ideal vs empirical

In this section the differences between *ideal* and *empirical* datasets are explained. An *ideal* dataset is the complete information of the timing of all events without any error. The *empirical* dataset is the *ideal* data, altered to represent what an empiricist would see when compiling a data set from reconstructed phylogenetic trees (i.e. no extinct species) to apply the DAISIE inference (maximum likelihood) model. In each island-mainland scenario the phylogenetic information of the island and mainland are shown. The assumptions of the DAISIE model are: there is no back-colonisation from the island to the mainland; cladogenetic speciation on the island or on the mainland produces two new species and the ancestor is no longer present (i.e. symmetric speciation); there is no divergence in the mainland species (i.e. genetically homogeneous) before they immigrate to the island; there is no incomplete lineage sorting so all species are correctly delineated in the tree; the timing of immigration events can be detected as the founder effect and evolution in isolation would leave a genetic signature. For the last assumption, the empirical data assumes that the phylogeny is complete at the population level, in order to determine divergence of populations of the same species on the island and mainland.

The DAISIE inference model needs information on the status of the clade (stac) in the following way:

stac 0: Empty island

stac 1: Non-endemic with unknown colonisation time but with a maximum to this colonisation time

stac 2: Endemic singleton or endemic clade

stac 3: Endemic singleton or endemic clade with one or more re-colonisations of the same mainland species

stac 4: Non-endemic singleton with known colonisation time

stac 5: Endemic singleton with unknown colonisation time, but with a maximum to this colonisation time

stac 6: Endemic clade with unknown colonisation time, but with a maximum to this colonisation time

stac 7: Endemic singleton or endemic clade and one or more re-colonisations of the same mainland species with unknown colonisation times, but with a maximum to this colonisation time

stac 8: Non-endemic with unknown colonisation time, but with a maximum and minimum to this colonisation time

stac 9: Endemic singleton with unknown colonisation time, but with a maximum and minimum to this colonisation time

In the *ideal* data only stac 2, 3 and 4 are assigned because it is known when the species colonises the island. However, in the *empirical* data stac 1, 5 and 6 can also be assigned when the colonisation time is not known but only a maximum (often island age). This is explained below.

Empty Island

If the island is empty at the end of the simulation, either through the failure of any mainland species to colonise, or if all the colonising species went extinct before the present then the *empirical* data is the same as the *ideal* data, which is that all the empiricist knows is that at the present there are no species. *Empirical* and *Ideal* data are assigned stac 0.

Single species on the island with no island events

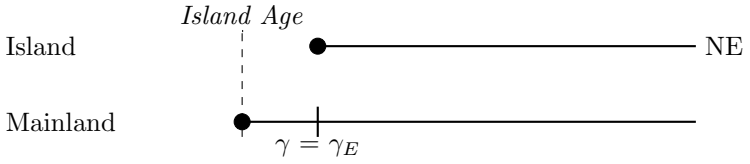

Figure A1: If the mainland species immigrates and does not go extinct or speciate, and no events happen on the island, the island species is non-endemic (NE), and the colonisation time in the *empirical* data (γ_E) is the same as the colonisation time in the *ideal* data (γ). *Empirical* and *ideal* data are assigned stac 4.

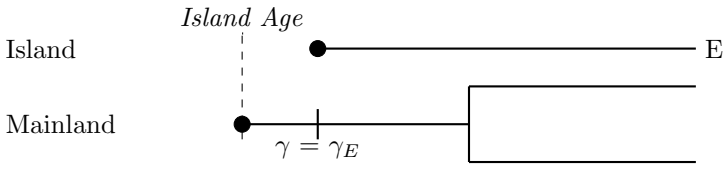

Figure A2: If the mainland species immigrates to the island then undergoes cladogenesis on the mainland and the descendent species do not go extinct and no events happen on the island, the island species is endemic (E) and in the *empirical* data the colonisation time (γ_E) is the same as the colonisation time in the *ideal* data (γ). *Empirical* and *ideal* data are assigned stac 2.

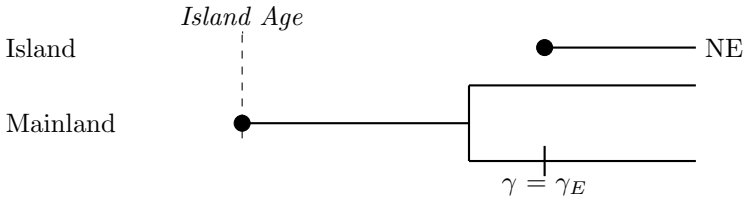

Figure A3: If the mainland species undergoes speciation and then one of the descendent species immigrates to the island and both the descendent species do not go extinct and no events happen on the island, the island species is non-endemic (NE) and in the *empirical* data the colonisation time (γ_E) is the same as the colonisation time in the *ideal* data (γ). *Empirical* and *ideal* data are assigned stac 4.

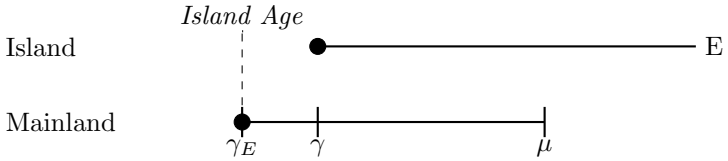

Figure A4: If the mainland species colonises the island and then goes extinct on the mainland (μ), without having speciated, and no events happen on the island, the island species is endemic (E) and in the *empirical* data the colonisation time (γ_E) is the maximum age of the island. *Ideal* data is assigned stac 2 and *empirical* data is assigned stac 5.

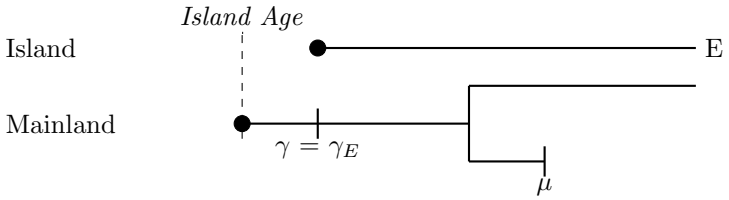

Figure A5: If the mainland species colonises the island and then undergoes speciation and one of the descendant species goes extinct (μ), and no events happen on the island, the island species is endemic (E) and in the *empirical* data the colonisation time (γ_E) is the same as the colonisation time in the *ideal* data (γ). *Ideal* and *empirical* data are assigned stac 2.

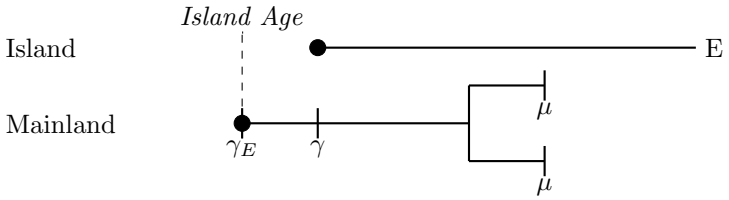

Figure A6: If the mainland species colonises the island and then undergoes speciation and both of the descendant species goes extinct (μ), and no events happen on the island, the island species is endemic (E) and in the *empirical* data the colonisation time (γ_E) is the maximum age of the island. *Ideal* data is assigned stac 2 and *empirical* data is assigned stac 5.

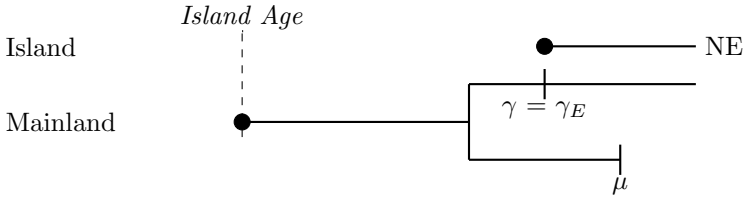

Figure A7: If the mainland species undergoes speciation and then one of the descendant species immigrates to the island and the other descendant goes extinct (μ) and no events happen on the island, the island species is non-endemic (NE) and in the *empirical* data the colonisation time (γ) is the same as the colonisation time in the *ideal* data (γ). *Ideal* and *empirical* data are assigned stac 4.

Figure A8: If the mainland species undergoes speciation and then one of the descendant species immigrates to the island and then goes extinct on the mainland (μ), and the other descendent survives, and no events happen on the island, the island species is endemic (E) and in the *empirical* data the colonisation time (γ_E) is the branching time on the mainland. *Ideal* and *empirical* data are assigned stac 2.

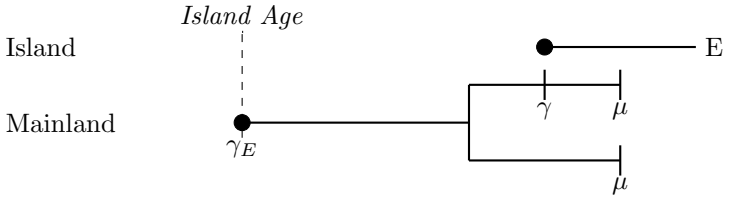

Figure A9: If the mainland species undergoes speciation and then one of the descendant species immigrates to the island and then both descendants go extinct (μ), and no events happen on the island, the island species is endemic (E) and in the *empirical* data the colonisation time (γ_E) is the maximum age of the island. *Ideal* data is assigned stac 2 and *empirical* data is assigned stac 5.

Single species on the island with anagenesis

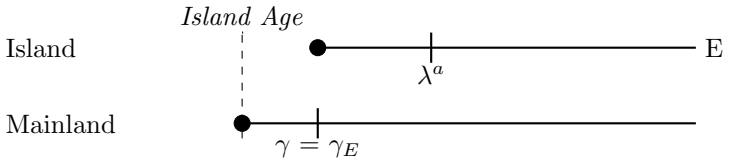

Figure A10: If the mainland species immigrates to the island and does not go extinct or speciate, and the island species undergoes anagenesis, the island species is endemic (E), and in the *empirical* data the colonisation time (γ_E) is the same as the colonisation time in the *ideal* data (γ). *Ideal* and *empirical* data is assigned stac 2.

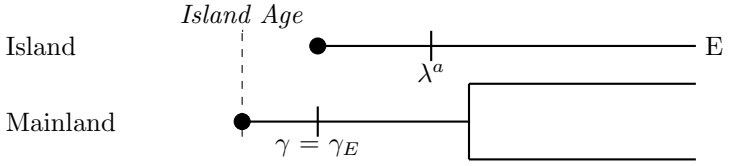

Figure A11: If the mainland species immigrates to the mainland and then undergoes speciation, with both mainland descendants surviving to the present, and the island species undergoes anagenesis, the island species is endemic (E) and in the *empirical* data the colonisation time (γ_E) is the same as the colonisation time in the *ideal* data (γ). *Ideal* and *empirical* data are assigned stac 2.

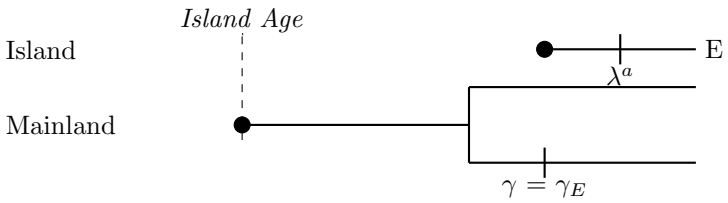

Figure A12: If the mainland species undergoes speciation and then one of the descendent species immigrates to the island and both the descendent species do not go extinct and the island species undergoes anagenesis, the island species is endemic (E) and in the *empirical* data the colonisation time (γ_E) is the same as the colonisation time in the *ideal* data (γ). *Ideal* and *empirical* data are assigned stac 2.

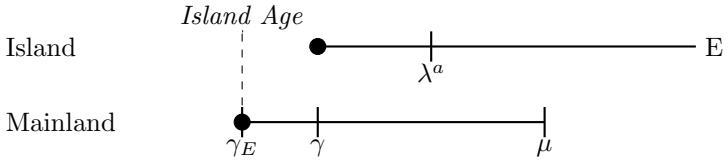

Figure A13: If the mainland species immigrates to the island (γ) and then goes extinct (μ) on the mainland, and the island species undergoes anagenesis, the island species is endemic (E), and in the *empirical* data the colonisation time (γ_E) is the maximum age of the island. *Ideal* data is assigned stac 2 and *empirical* data is assigned stac 5.

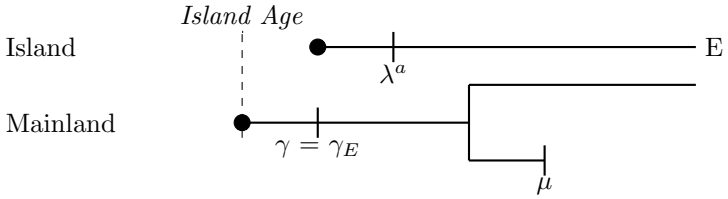

Figure A14: If the mainland species colonises the island (γ) and then undergoes speciation, and one or both of the descendent species go extinct on the mainland (μ), and the island species undergoes anagenesis, the island species is endemic (E), and in the *empirical* data the colonisation time (γ_E) is the same as the colonisation time in the *ideal* data (γ). *Ideal* and *empirical* is assigned stac 2.

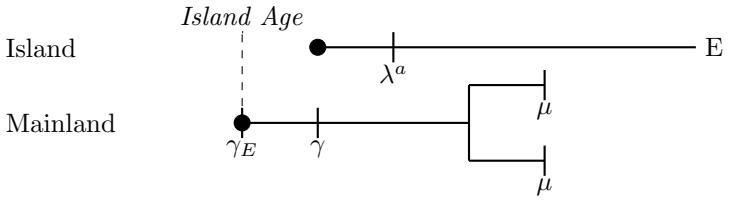

Figure A15: If the mainland species colonises the island and then undergoes speciation, and one or both of the descendent species go extinct (μ), and the island species undergoes anagenesis, the island species is endemic (E), and in the *empirical* data the colonisation time (γ_E) is the maximum age of the island. *Ideal* data is assigned stac 2 and *empirical* data is assigned stac 5.

Figure A16: If the mainland species undergoes speciation and then one of the descendant species immigrates to the island and the other descendant goes extinct (μ) and the island species undergoes anagenesis, the island species is endemic (E) and in the *empirical* data the colonisation time (γ_E) is the same as the colonisation time in the *ideal* data (γ). *Ideal* and *empirical* data are assigned stac 2.

Figure A17: If the mainland species undergoes speciation and then one of the descendant species immigrates to the island and then goes extinct on the mainland (μ), and the other descendent survives, and the island species undergoes anagenesis, the island species is endemic (E) and in the *empirical* data the colonisation time (γ_E) is the branching time on the mainland. *Ideal* and *empirical* are assigned stac 2.

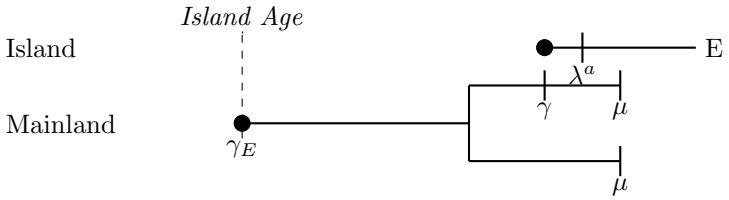

Figure A18: If the mainland species undergoes speciation and then one of the descendant species immigrates to the island (γ) and then both descendants go extinct on the mainland (μ), and the island species undergoes anagenesis, the island species is endemic (E) and in the *empirical* data the colonisation time (γ_E) is the maximum age of the island. *Ideal* data is assigned stac 2 and *empirical* data is assigned stac 5.

Clade on the island

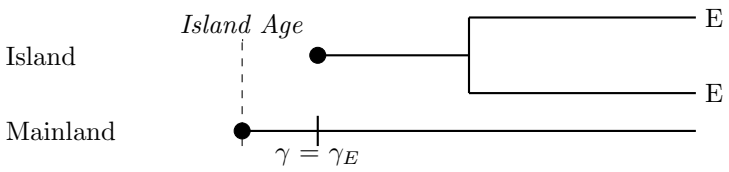

Figure A19: If the mainland species colonises the island and does not go extinct or speciate, and the island species undergoes cladogenesis the species are endemic (E), and in the *empirical* data the colonisation time (γ_E) is the same as the colonisation time in the *ideal* data (γ). *Ideal* and *empirical* data are assigned stac 2.

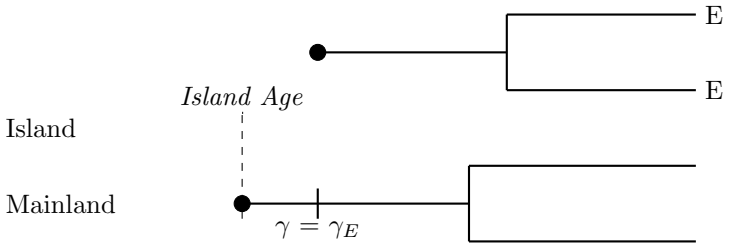

Figure A20: If the mainland species colonises the island and then undergoes speciation and neither of the descendent species go extinct and the species on the island undergoes cladogenesis both species are endemic (E) and in the *empirical* data the colonisation time (γ_E) is the same as the colonisation time in the *ideal* data (γ). *Ideal* and *empirical* data are assigned stac 2.

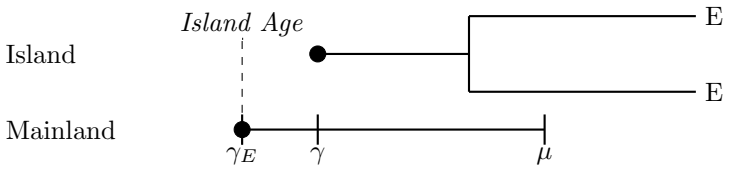

Figure A21: If the mainland species colonises the island (γ) and then goes extinct on the mainland (μ), and the island species undergoes cladogenesis the species are endemic (E), and in the *empirical* data the colonisation time (γ_E) is the maximum age of the island. *Ideal* data is assigned stac 2 and *empirical* data is assigned stac 6.

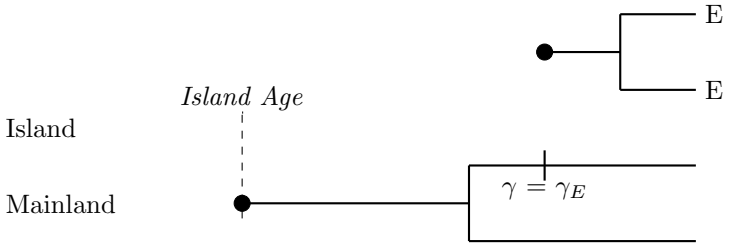

Figure A22: If the mainland species undergoes speciation and then one of the descendant species colonises the island and both descendent species survive to the present and the island species undergoes cladogenesis, the two island species are endemic (E), and in the *empirical* data the colonisation time (γ_E) is the same as the colonisation time in the *ideal* data (γ). *Ideal* and *empirical* data are assigned stac 2.

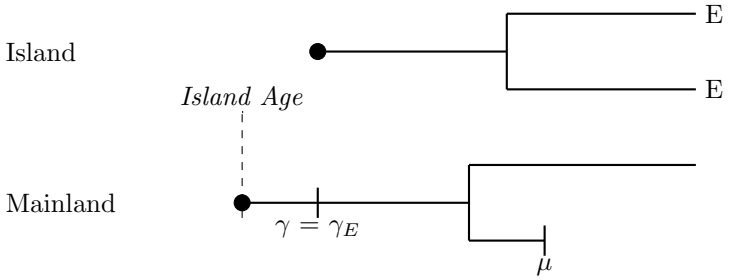

Figure A23: If the mainland species colonises the island, then undergoes speciation, and one of the descendent species goes extinct (μ) and the island species undergoes cladogenesis on the island, the two island species are endemic (E), and in the *empirical* data the colonisation time (γ_E) is the same as the colonisation time in the *ideal* data (γ). *Ideal* and *empirical* data are assigned stac 2.

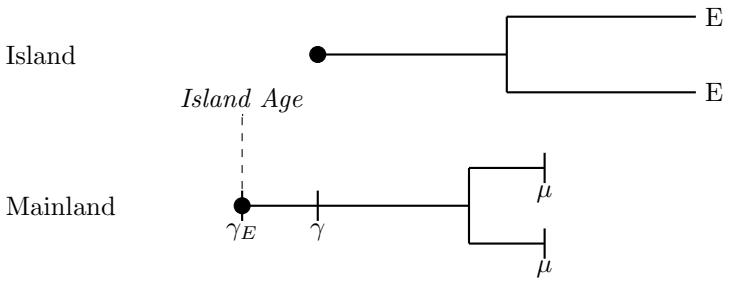

Figure A24: If the mainland species colonises the island (γ), then undergoes speciation, and both of the descendent species go extinct on the mainland (μ) and the island species undergoes cladogenesis on the island, the two island species are endemic (E), and in the *empirical* data the colonisation time (γ_E) is the maximum age of the island. *Ideal* data is assigned stac 2 and *empirical* data is assigned stac 6.

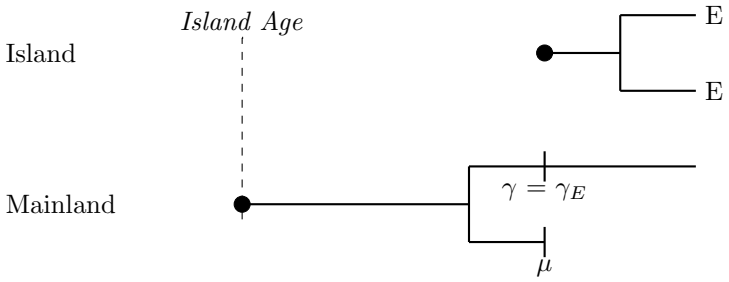

Figure A25: If the mainland species undergoes speciation, then colonises the island, and one of the descendent species goes extinct on the mainland (μ) and the island species undergoes cladogenesis on the island, the two island species are endemic (E), and in the *empirical* data the colonisation time (γ_E) is the same as the colonisation time in the *ideal* data (γ). *Ideal* and *empirical* data are assigned stac 2

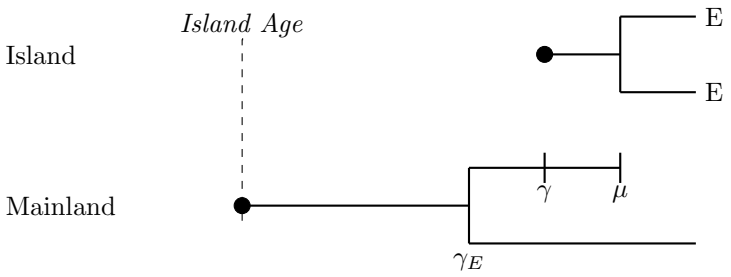

Figure A26: If the mainland species undergoes speciation, then one descendant colonises the island (γ), and then goes extinct on the mainland (μ) and the other descendant survives to the present, and the island species undergoes cladogenesis on the island, the two island species are endemic (E), and in the *empirical* data (γ_E) the colonisation time is the branching time on the mainland. *Ideal* and *empirical* data are assigned stac 2.

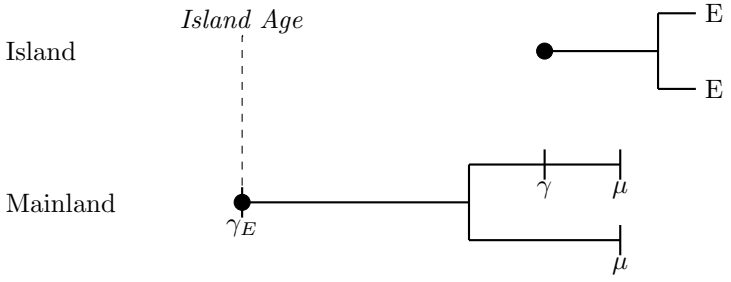

Figure A27: If the mainland species undergoes speciation, then colonises the island (γ), and both of the descendent species go extinct on the mainland (μ) and the island species undergoes cladogenesis on the island, the two island species are endemic (E), and in the *empirical* data the colonisation time (γ_E) is the maximum age of the island. *Ideal* data is assigned stac 2 and *empirical* data is assigned stac 6.

Colonisation of the same mainland species after cladogenesis or anagenesis on the island

In the case of multiple colonisations by the same mainland species, we assume that if the island species has not undergone cladogenesis or anagenesis and is thus still a non-endemic species, the new colonisation will overwrite the colonisation time of the first colonisation. The underlying reasoning is that gene flow from the new colonists will erase the genetic signature that the species was on the island before. If the island species has undergone cladogenesis or anagenesis and is thus an island endemic (or multiple island endemics in a clade), the mainland species can immigrate forming a non-endemic species on the island. In the case of a mainland branching event both descendent species are considered new species and so a re-immigration of the mainland ancestor cannot occur. Therefore, only scenarios where re-immigration without mainland speciation or re-immigration before a mainland speciation are considered. The effect of cladogenesis or anagenesis on re-colonisation is the same so only cladogenesis is shown in the section below, but all scenarios result in the same outcome in cases where anagenesis replaces the island branching time in the scenarios below.

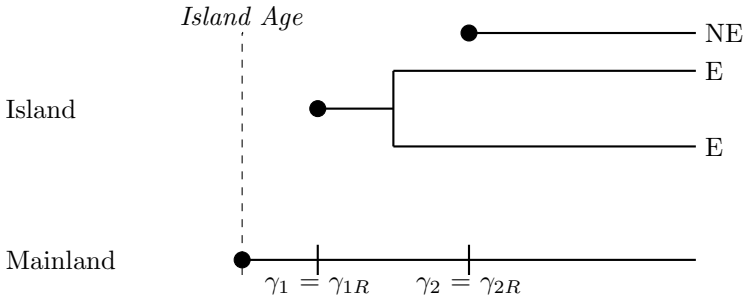

Figure A28: If the mainland species does not speciate or go extinct, and the mainland species colonises the island twice, the second colonisation taking place after the island species has undergone cladogenesis, the two species in the island clade are considered endemic (E), and the second immigrant is non-endemic (NE). The colonisation times in the *empirical* data (γ_{1E}) for the island clade is is the same as the colonisation time in the *ideal* data (γ_1) and for the non-endemic singleton the colonisation time in the *empirical* data (γ_{2E}) is the same as the colonisation time in the *ideal* data (γ_2). *Ideal* and *empirical* data are assigned stac 3.

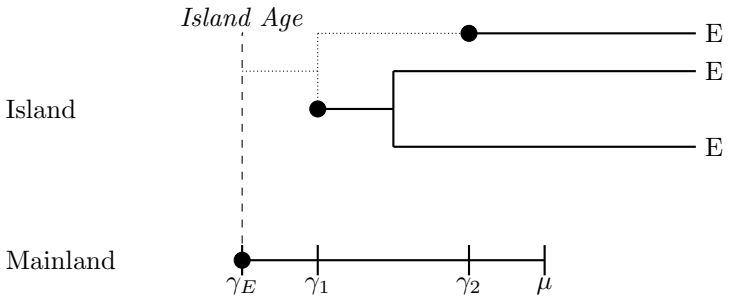

Figure A29: If the mainland species colonises the island twice and then goes extinct (μ) on the mainland, and between the immigration events the first island species underwent cladogenesis, all species on the island are considered endemic (E) and in the *empirical* data all species would be considered to arise from a single colonisation time (γ_E) at the maximum age of the island. Therefore the second colonisation time (γ_2) is lost, both colonists form a single clade (dotted line), and the singleton endemic is assumed to have colonised anywhere from the first immigration event. *Ideal* data is assigned stac 3 and *empirical* data is assigned stac 6.

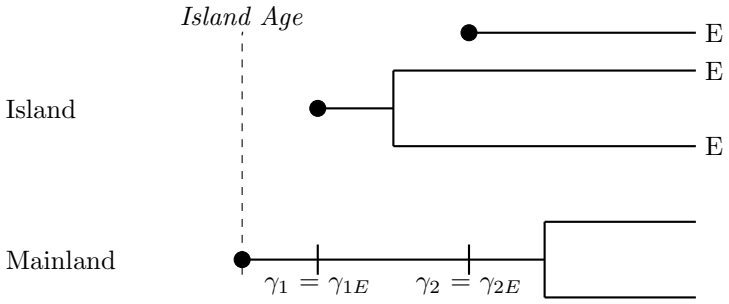

Figure A30: If the mainland species immigrates to the island twice and then undergoes a speciation event, with the island species having undergone cladogenesis between the immigration events, all species on the island are considered endemic (E), and in the *empirical* data the colonisation time (γ_{1E}) for the island clade is the same as the colonisation time in the *ideal* data (γ_1), and for the endemic singleton the *empirical* colonisation time (γ_{2E}) is the same as the colonisation time in the *ideal* data (γ_2), but these are not used in inference because *ideal* and *empirical* data are assigned stac 3.

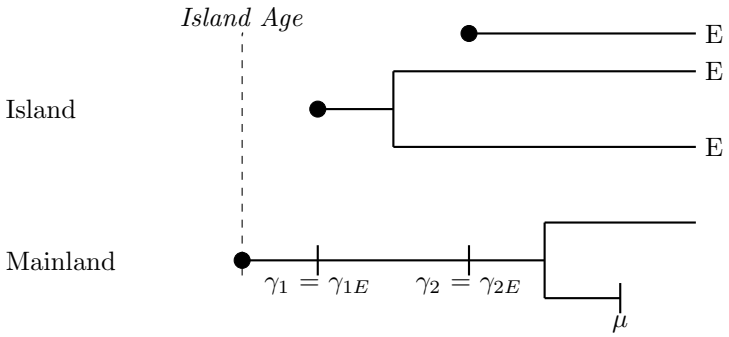

Figure A31: If the mainland species colonises the island twice and then speciates on the mainland, with one of the descendants going extinct (μ), and the first island species undergoing cladogenesis before the second immigration event, all species on the island are considered endemic (E), the colonisation times in the *empirical* data (γ_{1E}) for the island clade is the same as the colonisation time in the *ideal* data (γ_1) and for the endemic singleton the colonisation time in the *empirical* data (γ_{2E}) is the same as the colonisation time is the *ideal* data (γ_2), but these are not used in inference, because *ideal* and *empirical* data are assigned stac 3.

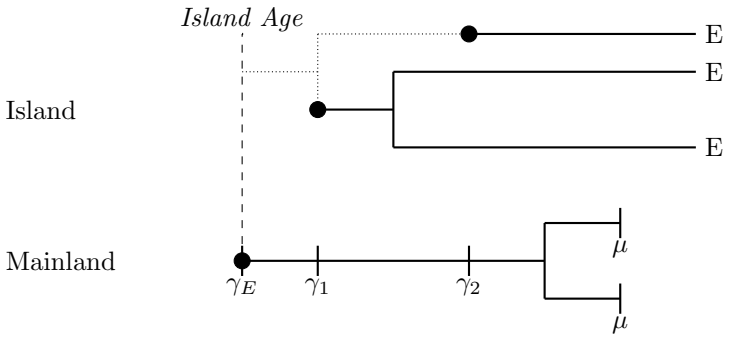

Figure A32: If the mainland species colonises the island twice and then speciates on the mainland, with both of the descendants going extinct (μ), and the first island species undergoing cladogenesis before the second immigration event, all species on the island are considered endemic (E) and in the *empirical* data all the species would be considered to arise from a single colonisation time (γ_E) at the maximum age of the island. Therefore the second colonisation time (γ_2) is lost and the singleton endemic is assumed to have colonised anywhere since the first immigration event (γ_1). *Ideal* data is assigned stac 3 and *empirical* data is assigned stac 6.

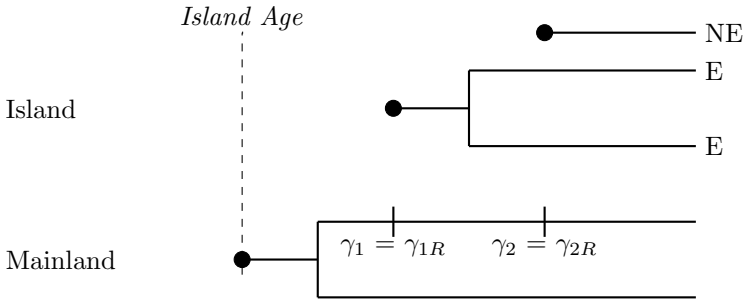

Figure A33: If the mainland species speciates and then one of the descendants colonises the island twice, with both of the descendants surviving, and the first island species undergoing cladogenesis before the second immigration event, the first colonist is an endemic clade (E) and the second colonist is a non-endemic singleton (NE) and the *empirical* is the same as the *ideal*. *Ideal* and *empirical* data are assigned stac 3, so the second colonisation time is lost with the first setting a maximum to the second.

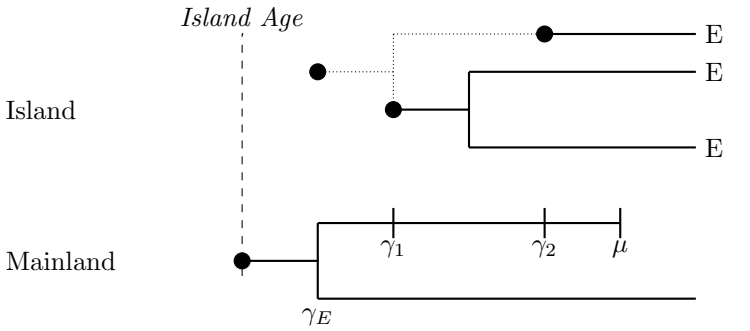

Figure A34: If the mainland species speciates and then one of the descendants colonises the island twice and then goes extinct (μ), and the other descendant survives, and the first island species undergoes cladogenesis before the second immigration event, both colonists are endemic (E). In the *empirical* case they are thought to form a single clade with a colonisation time at the branching time on the mainland (γ_E). The second colonisation time is lost. *Ideal* data is assigned stac 3 and *empirical* data are assigned stac 2.

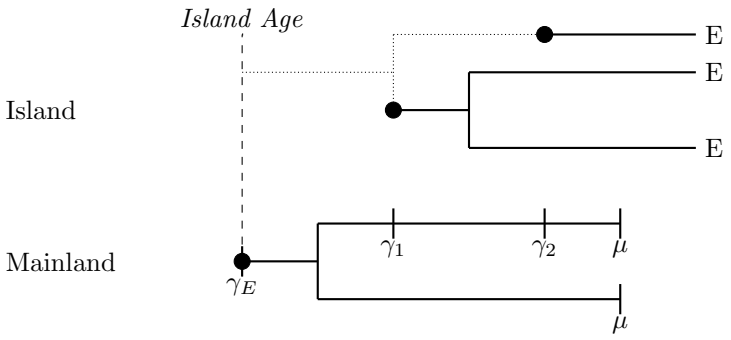

Figure A35: If the mainland species speciates and then one of the descendants colonises the island twice and then both descendants go extinct, and the first island species undergoing cladogenesis before the second immigration event, both colonists are endemic in the *empirical* case they are thought to form a single clade with a colonistation time at maximum island age. *Ideal* data is assigned stac 3 and *empirical* data are assigned stac 6.

Colonisation of a different mainland species after cladogenesis on the mainland with no island events

If the mainland species forms a clade, then different species can immigrate to the island that share the same mainland common ancestor. This is not considered in the DAISIE inference model, given the assumption that mainland species are single independent lineages that cannot undergo speciation or extinction. In these cases, when two or more different species immigrate to the island (as shown below), they evolve under the same diversity-dependent process. However, these are input into the DAISIE inference model as two island clades evolving under independent diversity-dependent processes. The effect of violating this assumption of the DAISIE inference model - along with the change in colonisation time - is one of the focal points of this study, and has empirical importance because mainland species are known to be phylogenetically non-independent, with some clades potentially sharing a recent common ancestor and thus likely to compete under the same diversity-dependent regime on the island. The scenarios below are generalisations of all scenarios of colonisations of different species from the same mainland clade, and can be extrapolated to any case with two or more colonisations of different species from the same clade.

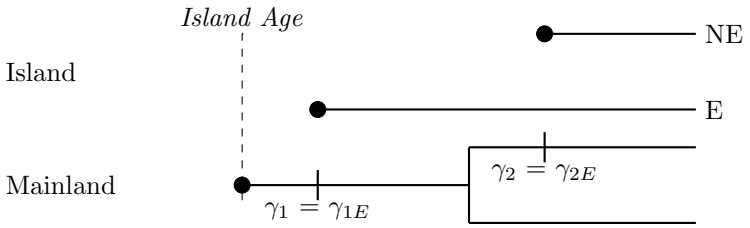

Figure A36: If the mainland species immigrates to the island, then undergoes speciation and one of the descendant species immigrates to the island and both descendants do not go extinct and no events happen on the island, the first island colonist is endemic (E), and the second island colonist is non-endemic (NE). In the *ideal* data the two immigration events are stored as two island clades. In the *empirical* data the first colonisation time (γ_{1E}) is the same as the colonisation time in the *ideal* data (γ_1), and the second colonisation time (γ_{2E}) is the same as the colonisation time in the *ideal* data (γ_2). In the *ideal* and *empirical* data the first colonist is assigned stac 2 and the second colonist is assigned stac 4.

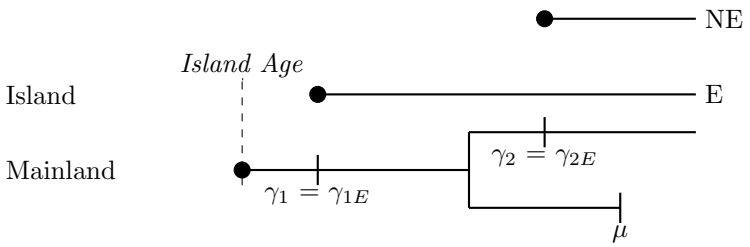

Figure A37: If the mainland species immigrates to the island, then undergoes speciation and one of the descendant species immigrates to the island and survives to the present on the mainland and the other descendant goes extinct and no events happen on the island, the first island colonist is endemic (E), and the second island colonist is non-endemic (NE). In the *ideal* data the two immigration events are stored as two island clades. In the *empirical* data the first colonisation time (γ_{1E}) is the same as the colonisation time in the *ideal* data (γ_1), and the second colonisation time (γ_{2E}) is the same as the colonisation time in the *ideal* data (γ_2). In the *ideal* and *empirical* data the first colonist is stored as stac 2 and the second colonist is stored as stac 4.

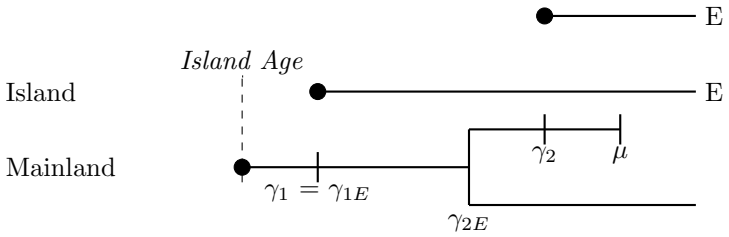

Figure A38: If the mainland species immigrates to the island, then undergoes speciation and one of the descendant species immigrates to the island and goes extinct on the mainland and the other descendant survives to the present, and no events happen on the island, both island colonists are endemic (E). In the *ideal* data the two immigration events are stored as two island clades. In the *empirical* data the first colonisation time (γ_{1E}) is the same as the colonisation time in the *ideal* data (γ_1), and the second colonisation time (γ_{2E}) is the branching time on the mainland. In the *ideal* and *empirical* data each colonist is stored as stac 2.

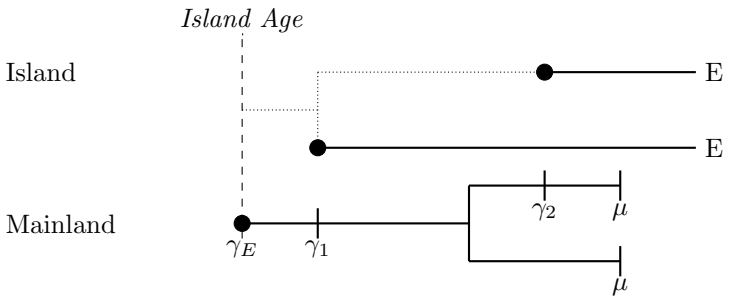

Figure A39: If the mainland species immigrates to the island, then undergoes speciation and one of the descendant species immigrates to the island and both descendants go extinct on the mainland, and no events happen on the island, both island colonists are endemic (E). In the *ideal* data the two immigration events are stored as two island clades. In the *empirical* data they are thought to for, a single clade with a colonisation time (γ_E) is the maximum age of the island. In the *ideal* data each colonist is stored as stac 2. In the *empirical* data the island clade is stored as stac 6.

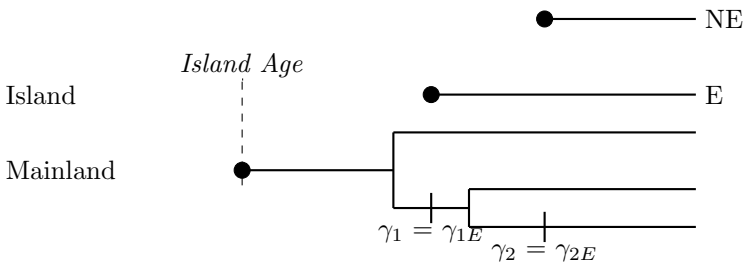

Figure A40: If the mainland species undergoes speciation and one descendant immigrates to the island, then undergoes speciation and one of those descendant species immigrates to the island and both descendants do not go extinct and no events happen on the island, the first island colonist is endemic (E), and the second island colonist is non-endemic (NE). In the *ideal* data the two immigration events are stored as two island clades. In the *empirical* data the first colonisation time (γ_{1E}) is the same as the colonisation time in the *ideal* data (γ_1), and the second colonisation time (γ_{2E}) is the same as the colonisation time in the *ideal* data (γ_2). In the *ideal* and *empirical* data the first colonist is stored as stac 2 and the second colonist is stored as stac 4.

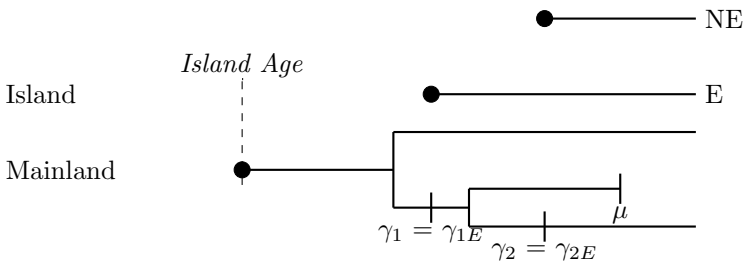

Figure A41: If the mainland species undergoes speciation and one descendant immigrates to the island, then undergoes speciation and one of those descendant species immigrates to the island and the other descendant goes extinct and no events happen on the island, the first island colonist is endemic (E), and the second island colonist is non-endemic (NE). In the *ideal* data the two immigration events are stored as two island clades. In the *empirical* data the first colonisation time (γ_{1E}) is the same as the colonisation time in the *ideal* data (γ_1), and the second colonisation time (γ_{2E}) is the same as the colonisation time in the *ideal* data (γ_2). In the *ideal* and *empirical* data the first colonist is stored as stac 2 and the second colonist is stored as stac 4.

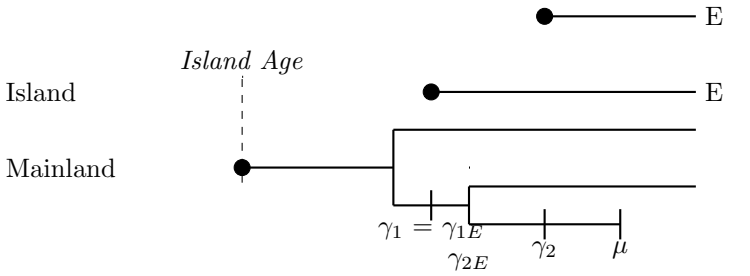

Figure A42: If the mainland species undergoes speciation and one descendant immigrates to the island, then undergoes speciation and one of those descendant species immigrates to the island then goes extinct and the other descendant survives to the present and no events happen on the island, both island colonists are endemic (E). In the *ideal* data the two immigration events are stored as two island clades. In the *empirical* data the first colonisation time (γ_{1E}) is the same as the colonisation time in the *ideal* data (γ_1), and the second colonisation time (γ_{2E}) is the last branching time on the mainland. In the *ideal* and *empirical* data each colonist is stored as stac 2.

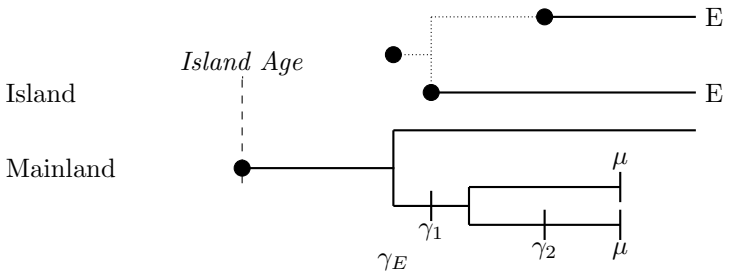

Figure A43: If the mainland species undergoes speciation and one descendant immigrates to the island then undergoes speciation and one of those descendant species immigrates to the island then both descendants go extinct and no events happen on the island, both island colonists are endemic (E). In the *ideal* data the two immigration events are stored as two island clades. In the *empirical* data they form a single clade both colonisation times (γ_{1E} and γ_{2E}) with a colonisation time at the first branching time on the mainland. In the *ideal* each colonist is stored as stac 2. In the *empirical* data the clade is stored as a stac 2.

Scenarios with more than one colonisation from different species from the same mainland clade with anagenesis or cladogenesis on the island can be determined based on the scenarios outlined above.