

Ecography

ECOG-03503

Valente, L., Phillimore, A. B. and Etienne, R. S. 2018.
Using molecular phylogenies in island biogeography:
it's about time. – Ecography doi: 10.1111/ecog.03503

Supplementary material

Appendix 1

Material and Methods

The biota of an island emerges from the species that are gained through colonisation and speciation, and those that are lost through extinction. DAISIE is a process-based stochastic dynamic model that allows these parameters to be estimated for complete insular communities (Valente et al. 2015). In this study, we treat the Galápagos archipelago as an ‘island’. We fitted a DAISIE model with four parameters (colonisation, cladogenesis, anagenesis and extinction) to each of the derived datasets and estimated parameters using maximum likelihood (Table A1). Note that a diversity-dependent version of this model with an additional parameter for the diversity carrying capacity is also available, but the diversity-dependent model was rejected for most of the Galápagos avifauna (Valente et al. 2015), and we therefore do not consider that model here. Maximum likelihood optimisations were conducted using the R package DAISIE v1.4. We used 40 initial random starting parameter values to avoid being trapped in local optima. We assumed a mainland pool size of 1000 species, and an archipelago age of 4 Myr. We did not condition on island occupation. In G1, the colonisation time of clades is not known, and, in such cases, the DAISIE inference method integrates over the possible colonisation times between the present and an upper bound, which in this case is the age of the archipelago. Therefore, all clades in G1 are assumed to have colonised at any point in time between the age of the island and the present. The Galápagos G1-G3 datasets are available in Dryad (pending).

To assess bias and precision of parameter estimates we conducted parametric bootstrap analyses (Fig. 1, A1 and Table A2). We assumed the ML parameters estimated under G3 are closest to the true values, and thus tested how well the model was able to recover the original parameters given different types of data. We simulated 3,000 datasets with the ML parameters of G3 and then created G1 and G2-type datasets from these by removing the relevant data elements. We then estimated DAISIE parameters from each of these datasets, from which we can infer precision and accuracy. Dispersion of the point cloud on Fig. 1 and S1

illustrates precision (the narrower the cloud, the higher precision); the difference between mean estimated values and the simulated value illustrates accuracy.

To check whether ML optimisations of the simulated datasets converge to the same point in parameter space, we performed a test on a subset of the simulated data. We ran optimisations with 10 random sets of initial starting values for each of 20 simulated datasets of G1, G2 and G3-type. We conducted this test for both small and large datasets (total 1,200 optimisations). For each of the G2 and G3 datasets all optimisations converged to the same likelihood and the same parameter set; therefore we are reasonably confident that we found the global optimum for all G2 and G3 datasets. Regarding G1, all optimisations per dataset converged to the same likelihood value, but in this case with very different parameter sets (as was the case for the empirical dataset). This means that the CIs we report for G1 are an underestimate.

Table A1 – Maximum likelihood parameters (biogeographical rates) estimated by fitting the DAISIE model to each of the three empirical datasets on the Galápagos avifauna. All rates are in number of events per lineage per million years. For G1 no optimum set of parameter values was found, therefore the range of parameters giving the same likelihood is shown.

Dataset	Colonisation	Extinction	Cladogenesis	Anagenesis
G1	0.006 - 0.49	1.42 - 139.58	1.44 - 120.00	0.51 - 19.76
G2	0.004	0.83	1	1.23
G3	0.009	2.68	2.55	1.01

Table A2 – Confidence intervals of parameter estimates obtained in the parametric bootstrap analyses. For each parameter, the 2.5–97.5 percentiles of the estimates across 3,000 simulated datasets are shown. Units are number of events per lineage per million years.

		Net diversification	Cladogenesis	Extinction	Colonisation	Anagenesis
Simulated value		-0.13	2.55	2.68	0.009	1.01
Simulation run for 4 million years	G1	-16.54 - 0.67	0.43 - 34.34	0 - 49.25	0.001 - 0.3	0 - 7.58
	G2	-1.75 - 0.61	0.67 - 5.8	0.63 - 6.72	0.002 - 0.03	0 - 7.3
	G3	-1.62 - 0.78	0.77 - 4.35	0.68 - 5.28	0.002 - 0.02	0 - 7.08
Simulation run for 10 million years	G1	-12.08 - 0.18	0.62 - 35.61	0.5 - 47.57	0.003 - 0.25	0 - 7.08
	G2	-1.03 - 0.05	1.14 - 6.14	1.51 - 6.49	0.004 - 0.03	0 - 7.05
	G3	-0.94 - 0.34	1.18 - 3.87	1.34 - 4.35	0.003 - 0.02	0 - 7.58

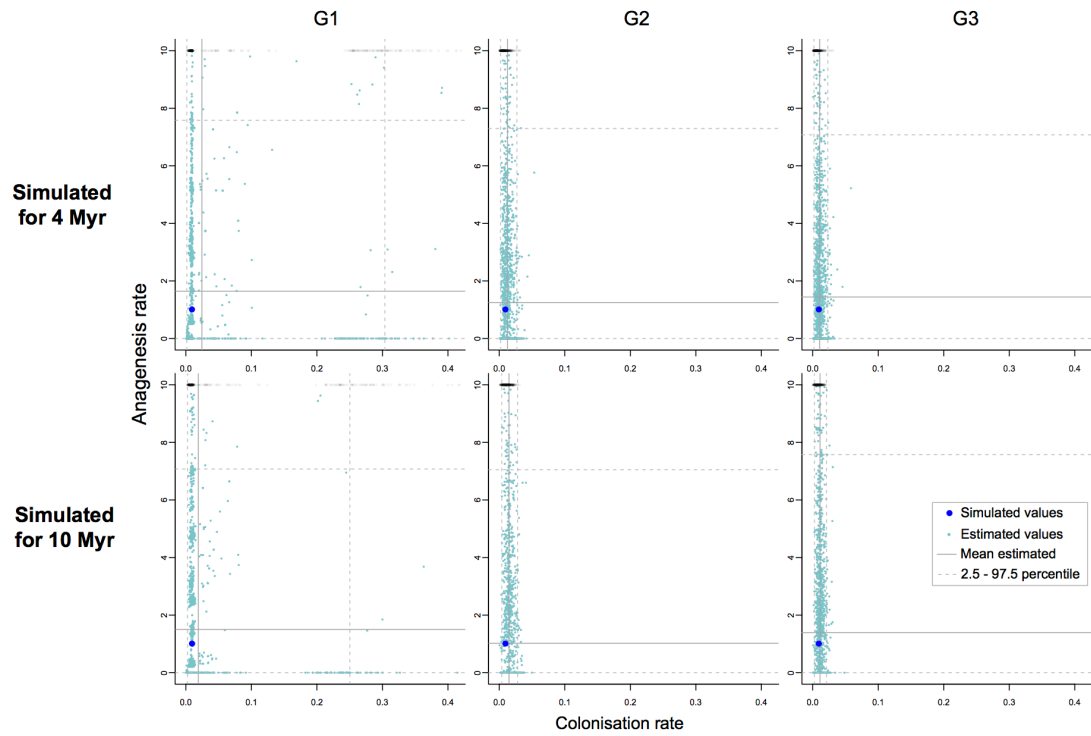


Figure A1 – Colonisation and anagenesis rate parameters estimated by fitting the DAISIE model to datasets with increasing amount of phylogenetic information. Rates are in events per lineage per million years. Black dots show cases where the anagenesis parameter was estimated as infinite, these are plotted with a value of 10 for visibility. Accuracy and precision are low for anagenesis for all datasets (G1-G3) and for both simulation lengths (4 and 10 Myr).