A simple dynamic model explains the diversity of island birds worldwide

A simple dynamic model explains island bird diversity worldwide

Luis Valente[1,2,3,4,*], Albert B. Phillimore[5], Martim Melo[6,7,8], Ben H. Warren[9], Sonya M. Clegg[10,11], Katja Havenstein[4], Ralph Tiedemann[4], Juan Carlos Illera[12], Christophe Thébaud[13], Tina Aschenbach[1], Rampal S. Etienne[3]

[1] Museum für Naturkunde, Leibniz Institute for Evolution and Biodiversity Science, Invalidenstraße 43, 10115 Berlin, Germany

[2] Naturalis Biodiversity Center, Understanding Evolution Group, Darwinweg 2, 2333 CR Leiden, the Netherlands

[3] Groningen Institute for Evolutionary Life Sciences, University of Groningen, P.O. Box 11103, Groningen 9700 CC, the Netherlands

[4] Unit of Evolutionary Biology/Systematic Zoology, Institute of Biochemistry and Biology, University of Potsdam, Karl-Liebknecht-Strasse 24–25, Haus 26, 14476 Potsdam, Germany

[5] Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, EH9 3JT, UK

[6] Museu de História Natural e da Ciência da Universidade do Porto, Praça Gomes Teixeira, 4099-002 Porto, Portugal

[7] CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, InBio, Laboratório Associado, Universidade do Porto, Campus Agrário de Vairão, 4485-661 Vairão, Portugal

[8] FitzPatrick Institute, DST-NRF Centre of Excellence, University of Cape Town, Rondebosch 7701, Cape Town, South Africa

[9] Institut de Systématique, Evolution, Biodiversité (ISYEB), Muséum National d'Histoire Naturelle, CNRS, Sorbonne Université, EPHE, UA, CP 51, 57 Rue Cuvier, 75005 Paris, France

[10] Edward Grey Institute, Department of Zoology, University of Oxford, Oxford, OX1 3PS, United Kingdom

[11] Environmental Futures Research Institute, Griffith University, Brisbane, Queensland, Australia, 4111

[12] Research Unit of Biodiversity (UO-CSIC-PA), Oviedo University, 33600 Mieres, Asturias, Spain

[13] Unité Mixte de Recherche 5174, CNRS-IRD-Paul Sabatier University, Toulouse, France

*Corresponding author. Email: luis.valente@naturalis.nl

Colonisation, speciation and extinction are dynamic processes that influence global patterns of species richness[1–6]. Island biogeography theory predicts that the contribution of these processes to the build-up of species diversity depends on area and isolation[7,8]. Remarkably, there has been no robust global test of this prediction[9], because neither the appropriate data nor the analytical tools have been available. Here, we address both deficiencies to reveal, for island birds, the empirical shape of the general relationships that determine how colonisation, extinction and speciation rates covary with island area and isolation. We compile the first global molecular phylogenetic dataset of birds on islands, based on the terrestrial avifaunas of 41 oceanic archipelagos worldwide (including 596 avian taxa), and apply novel methodology to estimate the sensitivity of island-specific rates of colonisation, speciation and extinction to island features (area, isolation). Our model predicts, with high explanatory power, several global relationships: a decline of colonisation with isolation, a decline of extinction with area, and an increase of speciation with area and isolation. Combining the theoretical foundations of island biogeography[7,8] with the temporal information contained in molecular phylogenies[10] proves a powerful approach to unveil the fundamental relationships that govern variation in biodiversity across the planet.

A key feature of global diversity is the tendency for some areas to harbour many more species than others[7,8]. Uncovering the drivers and regulators of spatial differences in diversity of simple systems such as islands is a crucial step towards understanding the global distribution of species richness. The two most prominent biodiversity patterns in fragmented or isolated environments worldwide are the increase of species richness with area and the decline of richness with isolation[8,11–14]. In their theory of island biogeography, MacArthur and Wilson proposed how the processes of colonisation and extinction could explain these patterns[7,8]. They argued that the rates of these processes are determined by the geographic context: colonisation decreases with isolation and extinction decreases with area[7,8]. They also suggested that rates of formation of island endemic species via *in situ* speciation increase with island isolation and area[8]. Despite an abundance of studies over five decades supporting the general patterns predicted by MacArthur and Wilson[2,15–18], tests of predictions regarding the dependence of the underlying processes – colonisation, speciation and extinction – on island geographic context (area and isolation) are few in number, and are either restricted in temporal, geographic, or taxonomic scope[5,19,20] or seek to infer speciation rates in the absence of data on the relationships among species[2,16]. As a result, there has been no robust and powerful test of MacArthur and Wilson's predictions on a global scale, and the effect of area and isolation on biogeographical processes acting on macro-evolutionary time scales remains largely unexplored.

Here we expand on approaches that leverage the information in time-calibrated molecular phylogenies of insular species[1,10,21,22] to determine how the processes of colonisation, speciation and extinction are influenced by area and isolation. The dynamic stochastic model DAISIE[10] (Dynamic Assembly of Islands through Speciation, Immigration, and Extinction) can accurately estimate maximum likelihood (ML) rates of colonisation, extinction and speciation rates (CES rates) from branching times (colonisation times and any *in situ* diversification events) and endemicity status of species resulting from one or multiple independent colonisations of a given island system (e.g. all native terrestrial birds on an archipelago)[10]. This method can also detect the presence or absence of diversity-dependence in rates of colonisation and speciation, by estimating a carrying capacity (upper bound to the number of species in an island system). Here we extend DAISIE to estimate, for the first time, the hyperparameters that

control the shape of the relationships between CES rates and the area and isolation of islands worldwide.

Accurate estimation of fundamental island biogeographic relationships requires suitable data from many archipelagos, but divergence-dated phylogenies of complete communities on islands remain scarce. Hence, we produced new dated molecular phylogenies for the terrestrial avifaunas of 41 archipelagos worldwide. By 'archipelago' we refer to both true archipelagos (composed of multiple islands) and isolated insular units consisting of single islands (e.g. Saint Helena). For each archipelago we compiled avian taxon lists (excluding introduced, marine, migratory, and aquatic species, as well as birds of prey, rails and nocturnal birds, see Methods) and collected physical data (Fig. 1, Supplementary Data 1 and 2). We use archipelagos as our insular unit, because the high dispersal abilities of birds within archipelagos imply that for birds, archipelagos can be considered equivalent to single islands for less dispersive taxa[23], and because archipelagos constitute the most appropriate spatiotemporal unit for framing analyses of biodiversity patterns at a large scale[2,24,25]. We extracted colonisation and speciation times for each archipelago from the phylogenetic trees, producing a 'global dataset' for the 41 archipelagos, which includes each archipelago's complete extant avifauna, plus all species known to have become extinct due to anthropogenic causes. The dataset comprises 596 insular taxa from 491 species. The phylogenies revealed a total of 502 archipelago colonisation events and 26 independent in-situ 'radiations' (cases where diversification has occurred within an archipelago) ranging in size from 2 to 33 species (the Hawaiian honeycreepers being the largest clade). The distribution of colonisation times is summarised in Fig. 1 and the full dataset is given in Supplementary Data 1.

Our extension of the DAISIE framework allows us to estimate hyperparameters that control the relationship between archipelago area and isolation and archipelago-specific local CES rates, i.e., rates of colonisation, cladogenesis (within-archipelago speciation involving *in situ* lineage splitting), anagenesis (within-archipelago speciation by divergence from the mainland without lineage splitting), natural extinction rates and carrying capacity. We tested the hypothesis that area and distance from the nearest mainland have an effect on the specific CES rates, and, where a significant effect was identified, estimated its shape and scaling. We developed a set of *a priori* models (Supplementary Table 1) where CES rates are power law functions of archipelago features. Area has been proposed to have a positive effect on cladogenesis and carrying

capacity[3,5,8], and a negative effect on extinction rates[8,26]. Archipelago isolation is hypothesised to reduce colonisation rates[7] and elevate anagenesis rates[27]. Models including or excluding diversity-dependence in rates of colonisation and cladogenesis[10] (i.e. estimating a carrying capacity parameter) were compared. We also considered a set of *post hoc* models with alternative shapes for the relationships (*post hoc* power and *post hoc* sigmoid models, see Methods, Supplementary Table 1).

We fitted a set of 28 candidate models to the global dataset using ML (Supplementary Table 2). The shape of the relationship of CES rates with area and distance for the two best models is shown in Fig. 2. Under the preferred *a priori* model (lowest value of Bayesian Information Criterion (BIC); M14, eight parameters) colonisation rates decline with archipelago isolation (exponent of the power law = –0.25 (95% confidence interval = –0.17 - –0.34)) and extinction rate decreases with area (scaling = –0.15 (–0.11 - –0.18)). Rates of cladogenesis increase with area (scaling = 0.26 (0.13 - 0.37)), whilst anagenesis increases with isolation (scaling = 0.42 (0.24 - 0.61)). The preferred *post hoc* model (M19, eight parameters) was also the preferred model overall and differs qualitatively from the preferred *a priori* model M14 only in the cladogenesis function. In M14 cladogenesis is solely a function of area, whereas in M19 cladogenesis depends interactively and positively on both area and distance from the nearest mainland, such that the cladogenesis-area relationship is steeper for more isolated archipelagos (Fig. 2 and Extended Data Fig. 1). In addition, we found no evidence for diversity-dependence, as the carrying capacity (K) was estimated to be much larger than the number of species on the island and models without a K parameter (no upper bound to diversity), such as M14 and M19, performed better than models including one (Supplementary Table 2). We also tested whether the inclusion of a combination of true archipelagos and single islands in our dataset could have affected our results, for example if opportunities for allopatric speciation are higher when an area is subdivided into multiple islands[28]. We repeated analyses excluding single island units and found that the same model (M19) is preferred with similar parameter estimates. Hence, we discuss only the results for the main dataset (including both single islands and true archipelagos). Our results are robust to uncertainty in colonisation and branching times (see section 'Sensitivity to alternative divergence times and tree topologies').

A parametric bootstrap analysis of the two preferred models (M14 and M19)

demonstrated that the method is able to recover hyperparameters with high precision and little bias (Extended Data Fig. 2). In order to test the significance of the relationships between area, isolation and CES rates, we conducted a randomization test on the global dataset with reshuffled areas and distances. This test estimated the exponent hyperparameters as zero in most reshuffled cases (i.e. no effect of area or isolation detected; Extended Data Fig. 3), confirming that it is the observed relationships between diversity and archipelago characteristics that generate our parameter estimates.

To assess model fit we simulated archipelago communities under the best model (M19) and found that for most archipelagos the observed diversity metrics (numbers of species, cladogenetic species and colonisations) were similar to the expected numbers, with some exceptions: for example, diversity was underestimated for Comoros and São Tomé & Príncipe (Fig. 3 and Extended Data Fig. 4). The ability of the model to explain observed values (pseudo-R^2 for total species = 0.72, cladogenetic species = 0.52, colonisers = 0.60) was very high considering the model includes only eight parameters (at least 12 parameters would be needed if each rate depended on area and isolation, and at least 164 parameters if each archipelago was allowed to have its own parameters) and was able to explain multiple diversity metrics. This represents a very large proportion of the explanatory power one would expect to obtain for data generated under the preferred model (Extended Data Fig. 5). Simulations under the best model reproduced the classic observed relationships between area, distance and diversity metrics (Fig. 4).

Our approach reveals the empirical shape of fundamental biogeographic relationships that have hitherto largely evaded estimation. In agreement with recent studies[2,29], we found strong evidence for a decline of rates of colonisation with isolation and of rates of extinction with area, confirming two of the key assumptions of island biogeography theory[7]. The colonisation-isolation effect was detected despite the decline of avian richness with distance from the nearest mainland in our empirical data not being as pronounced as in other less mobile taxa[4,11], revealing isolation to be a clear determinant of probability of immigration and successful establishment of populations even in a highly dispersive group such as birds. The extinction-area relationship has been a fundamental empirical generalization in conservation theory (for example for the design of protected areas[30]) but this is the first time the shape of this dependence is

characterized at the global spatial scale and macro-evolutionary time scale.

We provide novel insights into the scaling of speciation with area and isolation. Contrary to previous work on within-island speciation, which suggested the existence of an area below which cladogenesis does not take place on single islands[5], we do not find evidence for such an area threshold at the archipelago level, and under our model speciation is predicted to be non-zero even at small areas. In addition, our *post hoc* finding that rates of cladogenesis increase through an interactive effect of both island size and distance from the nearest mainland (Fig. 2 and Extended Data Fig. 1) provides a mechanism that limits radiations to archipelagos that are both large and remote[6,27]. Why this interaction exists requires further investigation, but one possibility is that unsaturated niche space provides greater opportunities for diversification[6]. In addition to the effects of physical features on cladogenesis, we found that rates of anagenesis increase with island isolation. While impressive insular radiations tend to receive the most attention from evolutionary biologists (e.g. Darwin's finches or Hawaiian honeycreepers), our phylogenies revealed that the majority of endemic birds in our dataset in fact display an anagenetic pattern (at the time of human arrival 231 of 350 endemic species had no extant sister taxa on the archipelago and there were only 26 extant in situ radiations). The positive effect of archipelago isolation on rates of anagenesis that we estimate suggests this fundamental but overlooked process is impeded by high levels of movement between island and mainland populations.

A variety of global patterns of biodiversity have been described – from small islands and lakes, up to biomes and continents - but the processes underpinning them remain little explored. Our simulations using parameters estimated from data were able to reproduce classic global patterns of island biogeography across 41 archipelagos (Fig. 4). This advances our understanding of macro-scale biology, by providing missing links between local process, environment and global patterns. Over half a century since the seminal work of MacArthur & Wilson[7], we now have the data and tools to go beyond statistical descriptions of diversity patterns, enabling us to quantify community-level processes that have long been elusive.

References

1. Ricklefs, R. E. & Bermingham, E. Nonequilibrium diversity dynamics of the Lesser Antillean avifauna. *Science.* **294**, 1522–1524 (2001).
2. Triantis, K. A., Economo, E. P., Guilhaumon, F. & Ricklefs, R. E. Diversity regulation at macro-scales: species richness on oceanic archipelagos. *Glob. Ecol. Biogeogr.* **24**, 594–605 (2015).
3. Whittaker, R. J., Triantis, K. A. & Ladle, R. J. A general dynamic theory of oceanic island biogeography. *J. Biogeogr.* **35**, 977–994 (2008).
4. Kreft, H., Jetz, W., Mutke, J., Kier, G. & Barthlott, W. Global diversity of island floras from a macroecological perspective. *Ecol. Lett.* **11**, 116–127 (2008).
5. Losos, J. B. & Schluter, D. Analysis of an evolutionary species-area relationship. *Nature* **408**, 847–850 (2000).
6. Gillespie, R. G. & Baldwin, B. G. Island biogeography of remote archipelagos. in *The Theory of Island Biogeography Revisited* (eds. Losos, J. & Ricklefs, R. E.) 358–387 (Princeton University Press, 2010).
7. MacArthur, R. H. & Wilson, E. O. An equilibrium theory of insular zoogeography. *Evolution (N. Y).* **17**, 373–387 (1963).
8. MacArthur, R. H. & Wilson, E. O. *The Theory of Island Biogeography*. (Princeton University Press, 1967).
9. Warren, B. H. *et al.* Islands as model systems in ecology and evolution: Prospects fifty years after MacArthur-Wilson. *Ecol. Lett.* **18**, 200–217 (2015).
10. Valente, L. M., Phillimore, A. B. & Etienne, R. S. Equilibrium and non-equilibrium dynamics simultaneously operate in the Galápagos islands. *Ecol. Lett.* **18**, 844–852 (2015).
11. Lomolino, M. V. Species-area and species-distance relationships of terrestrial mammals in the Thousand Island Region. *Oecologia* **54**, 72–75 (1982).
12. Diamond, J. M. Biogeographic kinetics: estimation of relaxation times for avifaunas of Southwest Pacific islands. *Proc. Natl. Acad. Sci.* **69**, 3199–3203 (1972).
13. Whittaker, R. J. & Fernandez-Palacios, J. M. *Island Biogeography: Ecology, Evolution, and Conservation*. (Oxford University Press, 2007).
14. Matthews, T. J., Rigal, F., Triantis, K. A. & Whittaker, R. J. A global model of island species–area relationships. *Proc. Natl. Acad. Sci.* 201818190 (2019).
15. Weigelt, P., Steinbauer, M. J., Cabral, J. S. & Kreft, H. Late Quaternary climate change shapes island biodiversity. *Nature* **532**, 99–102 (2016).
16. Lim, J. Y. & Marshall, C. R. The true tempo of evolutionary radiation and decline revealed on the Hawaiian archipelago. *Nature* **543**, 710–713 (2017).
17. Cabral, J. S., Weigelt, P., Kissling, W. D. & Kreft, H. Biogeographic, climatic and spatial drivers differentially affect α-, β-and γ-diversities on oceanic archipelagos. *Proc. R. Soc. London B Biol. Sci.* **281**, 20133246 (2014).
18. Matthews, T. J., Guilhaumon, F., Triantis, K. A., Borregaard, M. K. & Whittaker, R. J. On the form of species–area relationships in habitat islands and true islands. *Glob. Ecol. Biogeogr.* **25**, 847–858 (2016).
19. Simberloff, D. S. & Wilson, E. O. Experimental zoogeography of islands: the colonization of empty islands. *Ecology* **50**, 278–296 (1969).
20. Russell, G. J., Diamond, J. M., Reed, T. M. & Pimm, S. L. Breeding birds on small islands: island biogeography or optimal foraging? *J. Anim. Ecol.* **75**, 324–339 (2006).
21. Rabosky, D. L. & Glor, R. E. Equilibrium speciation dynamics in a model adaptive radiation of island lizards. *Proc. Natl. Acad. Sci.* **107**, 22178–22183 (2010).
22. Emerson, B. C. & Gillespie, R. G. Phylogenetic analysis of community assembly and structure over space and time. *Trends Ecol. Evol.* **23**, 619–30 (2008).
23. Kisel, Y. & Barraclough, T. G. Speciation has a spatial scale that depends on levels of gene flow. *Am Nat* **175**, 316–334 (2010).
24. Triantis, K., Whittaker, R. J., Fernández-Palacios, J. M. & Geist, D. J. Oceanic archipelagos: a perspective on the geodynamics and biogeography of the World's smallest biotic provinces. *Front. Biogeogr.* **8**, e29605 (2016).

25. Santos, A. M. C. *et al.* Are species–area relationships from entire archipelagos congruent with those of their constituent islands? *Glob. Ecol. Biogeogr.* **19**, 527–540 (2010).
26. Ricklefs, R. E. & Lovette, I. J. The roles of island area per se and habitat diversity in the species-area relationships of four Lesser Antillean faunal groups. *J. Anim. Ecol.* **68**, 1142–1160 (1999).
27. Rosindell, J. & Phillimore, A. B. A unified model of island biogeography sheds light on the zone of radiation. *Ecol. Lett.* **14**, 552–560 (2011).
28. Losos, J. B. & Ricklefs, R. E. Adaptation and diversification on islands. *Nature* **457**, 830–836 (2009).
29. Keil, P. *et al.* Extinction rate has a complex and non-linear relationship with area. *bioRxiv* 81489 (2016). doi:10.1101/081489
30. Wilcox, B. A. & Murphy, D. D. Conservation strategy: the effects of fragmentation on extinction. *Am. Nat.* **125**, 879–887 (1985).

MAIN TEXT FIGURE LEGENDS

Figure 1 – Archipelago and island bird colonisation time data. Circles show number of species belonging to our focal group (both extinct and extant) found in each archipelago (at the time of human arrival). Numbers on map correspond to numbers to the left of the archipelago name. Numbers to the right of the archipelago name: number of species from our focal assemblage on the archipelago | percentage of species sampled in the phylogenetic trees. Even species not sampled in the trees are accounted for by including them as missing species that could have colonised any time since emergence of the archipelago. Colonisation times plot: grey horizontal lines – archipelago ages (Extended Data Table 1). Violin plots (blue) show the kernel density of the distribution of times of colonisation of bird species in each archipelago, obtained from the phylogenetic trees. Thick black line inside violin - interquartile distance; thin black line - 95% CI; black dot - median. Archipelagos with no violin plot or dots are cases for which no species of our focal assemblage were present at the time of human arrival, or none were sampled using molecular data. Birds from left to right: Seychelles sunbird, Seychelles magpie robin, silvereye, Príncipe thrush, laurel pigeon, dodo (extinct), Mauritius fody, red-moustached fruit dove (extinct), Galápagos warbler, Norfolk kaka (extinct). Bird images used with permission from: Cláudia Baeta, Pedro Cascão, Martijn Hammers, Julian Hume, Dubi Shapiro and Juan Varela. There are no *in-situ* radiations in the Mascarenes (Mauritius, Reunion and Rodrigues) because we treat the islands as separate entities (but see sensitivity analyses).

Figure 2 – Estimated relationships between island area and isolation and local island biogeography parameters. Isolation measured as distance to the nearest mainland (D_m). Based on the maximum likelihood global hyperparameters of the best models (equations describing the relationships given in Supplementary Table 1). Darker lines – M14 model, lighter lines – M19 model. Under the M14 model, cladogenesis rate depends only on area. Under the M19 model, cladogenesis rate increases with both area and D_m, and thus lines for more (far, 5,000 km) and less (near, 50 km) isolated islands are shown. See also Extended Data Fig. 1 for the relationship of cladogenesis with both area and distance under the M19 model.

Figure 3 – Goodness of fit of the preferred model (M19). The map identifies whether the diversity metrics were well estimated (empirical value matches 95% confidence interval of simulations), underestimated (empirical value higher than 95% interval) or overestimated (empirical value lower than 95% interval). Intervals based on 1000 simulations of each archipelago (see Extended Data Fig. 4). Numbers indicating archipelagos on the map match those in Fig. 1.

Figure 4 – Observed and predicted island diversity-area and island diversity-distance relationships. Grey vertical lines show the 95% confidence intervals across 1,000 datasets simulated for each of the 41 archipelagos assuming the M19 model. Blue points: mean values of the simulations; blue line – fitted line for the simulated data; red points – observed values in the empirical data; red line – fitted line for the empirical data; red shaded area is the 95% confidence interval of the predicted relationship for the empirical data.

Methods

Archipelago selection

We focus on oceanic islands, i.e. volcanic islands that have never been connected to any other landmass in the past. We also include the Granitic Inner Seychelles, even though these islands have a continental origin, because they have been separated from other landmasses for a very long period of time (64 million years (Ma)[31]) and can be considered quasi-oceanic, as all extant avian species originated in much more recent times. The 41 archipelagos chosen are located in the Atlantic, Indian and Pacific oceans, with latitudes between 45° North and South. Islands within these archipelagos are separated by a maximum of 150 km. The sole exceptions are the Azores and Hawaii, two very isolated systems where the distances between some islands exceed this value. The shape files used to plot the maps of Figs. 1 and 3 were obtained from Weigelt, Jetz and Kreft 2013[32].

Physical and geological data

Full archipelago data is given in Supplementary Data 2 and Extended Data Table 1. We obtained data on total contemporary landmass area for each archipelago. For our isolation metric, we computed the minimum round earth distance to the nearest mainland (D_m) in km using Google Earth. We considered 'nearest mainland' to be the nearest probable source of colonists (but see 'Sensitivity to archipelago selection and isolation metrics' section for different isolation metrics). This is the nearest continent except for island groups that were closer to Madagascar, New Guinea or New Zealand than to the continent, in which case we assigned these large continent-like islands as the mainland. This is supported by our phylogenetic data – for example, many Indian Ocean island taxa have closest relatives on Madagascar rather than mainland Africa.

Island palaeo-areas and past archipelago configurations have been shown to be better predictors of endemic insular diversity than contemporary area[15,33]. In contrast, island total native and non-endemic richness is better predicted by present island characteristics[15,33]. With insufficient data on island ontogeny being available (i.e. describing empirical area trajectories from island birth to present) we therefore

analysed contemporary area and isolation as currently the most appropriate units for our dataset.

We conducted an extensive survey of the literature and consulted geologists to obtain archipelago geological ages (Extended Data Table 1), treating the age of the oldest currently emerged island as an upper bound for colonisation. Islands may have been submerged and emerged multiple times and we consider the age of the last known emergence. For the Aldabra Group we used an age older than the published estimate. The current estimated age of re-emergence of Aldabra is 0.125 Ma[34], but nine out of 12 Aldabra colonisation events in our dataset are older, suggesting the archipelago was not fully submerged prior to this and may have been available for colonisation for a longer period. Therefore, for Aldabra we used an older upper bound of 1 Ma for colonisation, although we acknowledge that the mitochondrial markers used for dating may not provide sufficient resolution at the shallow temporal scale of the published age. For Hawaii, the colonisation times we obtained for more than half of the colonisation events were older than the age of the current high islands that is often used as a maximum age for colonisation (~5 Ma). Therefore, instead of this age, we used the much older estimate of 29.8 Ma of the Kure Atoll[35] to account for currently submerged or very low-lying Hawaiian islands that could have received colonists in the past. For Bermuda and Marianas, we could not find age estimates in the literature, and we therefore consulted geologists to obtain these (P. Hearty, R. Stern and M. Reagan, pers. comm., Extended Data Table 1).

Island avifaunas

Our sampling focused on native resident terrestrial birds and we considered only birds that colonise by chance events (e.g. hurricanes, rafts). We thus excluded marine and migratory species, because they are capable of actively colonising an island at a much higher rate. We focused on songbird-like and pigeon-like birds, which constitute the majority of terrestrial (land-dwelling) birds on islands. Following a precedent set by previous work[10,27,36], we included only species from the same trophic level (in the spirit of MacArthur and Wilson's model): we excluded aquatic birds, birds of prey, rails (many are flightless or semi-aquatic) and nightjars (nocturnal). We also excluded introduced and vagrant species. Including species such as rails and owls (which are components of

many island avifaunas) would have led to a higher estimate of the product of colonisation rate and mainland pool size due to a larger mainland pool, and potentially to higher estimated rates of anagenesis (many owl or rail species are island endemics with no close relatives on the islands).

For the focal avian groups, we compiled complete taxon lists for each of the 41 archipelagos based on recent checklists from Avibase (http://avibase.bsc-eoc.org), which we cross-checked with the online version of the *Handbook of the Birds of the World* (HBW[37]). We followed HBW's nomenclature and species assignations, except for 12 cases where our phylogenetic data disagree with HBW's scheme (noted in the column 'Taxonomy' of Supplementary Data 1). For example, in eleven cases phylogenetic trees support raising endemic island subspecies to species status (we sampled multiple samples per island taxon and outgroup, and the island individuals form a reciprocally monophyletic well-supported clade), and for these taxa we decided it was more appropriate to use a phylogenetic species concept so as not to underestimate endemicity and rates of speciation (Supplementary Data 1). We re-ran DAISIE analyses using HBW's classification and found that the ML parameters are very similar and thus we report only the results using the scheme based on the phylogenies produced for this study.

For each bird species found on each archipelago we aimed to sample sequence data for individuals on the archipelago and the closest relatives outside the archipelago (outgroup taxa). Our sampling success per archipelago is shown on Fig. 1 and Extended Data Table 1.

Extinct species

We do not count extinctions with anthropogenic causes as impacting the natural background rate of extinction. Therefore, we explicitly include species where there is strong evidence that they have been extirpated by humans. We treat taxa extirpated on an archipelago by humans as though they had survived in that archipelago until the present following the approach of Valente, Etienne and Dávalos 2017[38].

We identified anthropogenic extinctions based on published data[39–46] and personal comments (Josep Antoni Alcover and Juan Carlos Rando on unpublished Macaronesian taxa; Ferran Sayol and Søren Faurby). We include the species present on

the islands that belong to our archipelago definition as in Supplementary Data 2. We excluded largely hypothetical accounts or pre-Holocene fossils that greatly pre-date human arrival. Our dataset accounts for 153 taxa that were present upon first human contact and have gone extinct since, probably because of human activities including human introduction of invasive species. To our knowledge 71 of these taxa have previously been sequenced using ancient DNA or belong to clades present in our trees, and we were thus able to include them in the phylogenetic analyses as regular data (n = 54), or as missing species by adding them as unsampled species to a designated clade (n = 17). For the remaining 82 extinct taxa, sequences were not available and we were unable to obtain samples and to allocate them to clades. We assume that these taxa represent extinct independent colonisations and we included them in the analyses using the "Endemic_MaxAge" and "Non_endemic_MaxAge" options in DAISIE, which assume that they have colonised at any given time since the birth of the archipelago (but before any *in situ* cladogenesis event). As an example, our dataset includes the 27 species of Hawaiian birds belonging to our focal group that are known to have gone extinct since human colonisation. Eight of these species were included using DNA data, 17 were added as missing species to their clades (14 honeycreepers and 3 *Myadestes*) and two were added using the Endemic_MaxAge option in DAISIE (*Corvus impluviatus* and *Corvus viriosus*).

Sequence data: GenBank

We conducted an extensive search of GenBank for available DNA sequences from the 596 island bird taxa fitting our sampling criteria and from multiple outgroup taxa, using software Geneious 11[47]. The molecular markers chosen varied from species to species, depending on which marker was typically sequenced for the taxon in question, the commonest being cytochrome b (cyt-*b*). In total, we downloaded 3155 sequences from GenBank. For some taxa, sequences from both archipelago and close relatives from outside the archipelago were already available from detailed phylogenetic or phylogeographic analyses. In some cases, a target species had been sampled, but only from populations outside the archipelago. In other cases, the species on the archipelago had been sampled, but the sampling of the relatives outside of the archipelago was lacking or only from distant regions, which meant a suitable outgroup was not available

on GenBank. Finally, for some species there were no previous sequences available on GenBank. GenBank accession numbers and geographical origin for the downloaded sequences are provided on the DNA matrices and maximum clade credibility trees (uploaded to Mendeley Data).

Sequence data: new samples

Sequences available on GenBank covered only 54% (269/502) of the total independent colonisation events. We improved the sampling by obtaining new sequences for many island taxa (n = 174 taxa) and from their close relatives from continental regions (n = 78). We obtained new samples from three sources: field trips, research collections and colleagues who kindly contributed field samples. New samples were obtained during field trips conducted by M.M. (Gulf of Guinea and African continent); B.H.W. and C.T. (Comoros, Mauritius Isl., Rodrigues, Seychelles); S.M.C. (New Caledonia); J.C.I. (Macaronesia, Europe and Africa) and L.V. (New Caledonia), between 1999 and 2017. Samples of individuals were captured using mist-nets or spring traps baited with larvae. Blood samples were taken by brachial venipuncture, diluted in ethanol or Queen's lysis buffer in a microfuge tube. Birds were released at the point of capture. Aldabra Group samples were obtained from research collections of the Seychelles Islands Foundation. Museum samples from several Galápagos and Comoros specimens were obtained on loan from respectively, the California Academy of Sciences and the Natural History Museum London. Additional samples from various localities (Aldabra Islands, Iberian Peninsula, Madagascar and Senegal) were kindly provided by collaborators, as indicated in Supplementary Table 3. Sample information and GenBank accession numbers for all new specimens are provided in Supplementary Table 3.

DNA was extracted from blood, feathers and museum toe-pad samples using QIAGEN DNeasy Blood and Tissue kits (QIAGEN, USA). For museum samples, we used a dedicated ancient DNA lab facility at the University of Potsdam to avoid contamination. The cyt-*b* region (1100 base pairs) was amplified using the primers shown in Extended Data Table 2. DNA from historical museum samples was degraded and cyt-*b* could not be amplified as a single fragment. We thus designed internal primers to sequence different overlapping fragments in a stepwise fashion (Extended Data Table 2).

Polymerase chain reactions (PCR) were set up in 25 µL total volumes including 5

μL of buffer Bioline MyTaq, 1 μL (10 mM) of each primer, and 0.12 μL MyTaq polymerase. PCRs were performed with the following thermocycler conditions: initial denaturation at 95° C for 1 min followed by 35 cycles of denaturation at 95° C for 20 s, with an annealing temperature of 48° C for 20 s, and extension at 72° C for 15 s min and a final extension at 72° C for 10 min. Amplified products were purified using Exonuclease I and Antarctic Phosphatase, and sequenced at the University of Potsdam (Unit of Evolutionary Biology/Systematic Zoology) on an ABI PRISM 3130xl sequencer (Applied Biosystems) using the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems). We used Geneious 11 to edit chromatograms and align sequences.

Phylogenetic analyses

To estimate times of colonisation and speciation for each archipelago we produced new divergence dated phylogenies or compiled published dated trees, to yield a total of 91 independent phylogenies (maximum clade credibility trees and posterior distribution deposited in Mendeley for all new trees produced for this study; the 11 previously published trees are available upon request). Information on all alignments and trees, including molecular markers, source of data, calibration method and substitution model are given in Extended Data Tables 3 and 4 and Supplementary Table 4. The majority of alignments/phylogenies focus on a single genus, but some include multiple closely related genera or higher order clades (family, order) depending on the diversity and level of sampling of the relevant group (taxonomic scope indicated in Extended Data Tables 3 and 4). Most alignments include taxa from a variety of archipelagos. Alignments were based on a variety of markers, according to which marker had been mostly sequenced for a given group.

For the new dating analyses conducted for this study, we created 80 separate alignments for different groups using a combination of sequences from GenBank (n = 3155) and new sequences (n = 252) produced for this study. In some cases, we obtained DNA alignments directly from authors of previous studies and these are credited in Extended Data Table 3. Phylogenetic divergence dating analyses were performed in BEAST 2[48]. For each alignment we performed substitution model selection in jModeltest[49] using the Bayesian information criterion. We used rates of molecular

evolution for avian mitochondrial sequences, which have been shown to evolve in a clock-like fashion at an average rate of ~2% per Ma[50]. Molecular rate calibrations can be problematic for ancient clades, due to high levels of heterotachy in birds[51]. In addition, mitochondrial DNA saturates after about 10 to 20 million years, and genetic distances of more than 20% may provide limited information regarding dating[52]. Therefore, we only used molecular rate dating to extract node ages for branching events at the tips of the trees, at the species or population level (oldest colonisation time in our dataset is 15.3 Ma, but most are much younger). Rates of evolution were obtained from the literature and varied between different markers and taxonomic group (Supplementary Table 4). We applied the avian mitochondrial rates estimated from cyt-*b* by Weir and Schluter[50] (but see 'Sensitivity to alternative divergence times and tree topologies' section for different rates).

We applied a Bayesian uncorrelated lognormal relaxed clock model. For each analysis, we ran two independent chains of between 10 and 40 million generations, with a birth-death tree prior. We assessed convergence of chains and appropriate burn-ins with Tracer, combined runs using LogCombiner, and produced maximum clade credibility trees with mean node heights in Tree Annotator. We produced a total of 80 maximum clade credibility trees.

For 11 groups (Extended Data Table 4), well-sampled and rigorously-dated phylogenies were already available from recent publications, all of which conducted Bayesian divergence dating using a variety of calibration methods, including fossils and molecular rates. We obtained maximum clade credibility trees from these studies from online repositories or directly from the authors (Extended Data Table 4).

Colonisation and branching times

The nodes selected in the dated trees for estimates of colonisation and branching times are given for each taxon in Supplementary Data 1. Our node selection approach was as follows. For cases in which samples representing species or populations from archipelagos formed a monophyletic clade consisting exclusively of archipelago individuals, we used the stem age of this clade as colonisation time. For cases in which only one individual of the archipelago was sampled, we used the length of the tip leading to that individual, which is equivalent to the stem age. For cases in which the archipelago individuals were embedded in a clade containing mainland individuals of

the same species, i.e. paraphyly or polyphyly; we assumed (based on morphological characteristics) that this is due to incomplete lineage sorting of the insular and mainland lineages, and we therefore used the MRCA of the archipelago individuals, or the crown node when the MRCA coincides with the crown. For these later cases using the stem would most likely have been an overestimation of the colonisation time, as we assume that colonisation happens from the mainland to the archipelago. For such cases we applied the ages using the "MaxAge'' option in DAISIE, which integrates over the possible colonisation times between the present and the upper bound. A robustness test of our results to node choice is given in section "Sensitivity to alternative branching times and tree topologies".

For a total of 19 endemic taxa we could not obtain sequences, but we could allocate them to a specific island clade (e.g. Hawaiian honeycreepers and solitaires). These were added as missing species to that clade. For 96 non-endemic taxa we could not obtain sequences of individuals from the archipelago, but we could obtain sequences from the same species from different regions. For these cases we used the crown or the stem age of the species as an upper bound for the age of the colonisation event, using the ''Non_endemic_MaxAge'' option in DAISIE. Finally, for 124 taxa (20.8 %) no sequences of individuals from the archipelago were available on GenBank and we were not able to obtain samples for sequencing from the species or from close relatives. We assumed these cases constituted independent colonisations that could have taken place any time since the origin of the archipelago and the present, and applied the ''Non_endemic_MaxAge'' and "Endemic_MaxAge" options in DAISIE with a maximum age equal to the archipelago age. DAISIE makes use of this information[53].

Global dataset characteristics

Data points from taxa of the same archipelago were assembled into 41 archipelago-specific datasets. These 41 datasets were in turn assembled into a single dataset (D1) which was analysed with DAISIE (D1 DAISIE R object, available in Mendeley Data https://doi.org/10.17632/sy58zbv3s2.2). This dataset (information Supplementary Data 1) has a total of 596 taxa (independent colonisation events plus species within radiations), covering 491 species from 203 different genera and eight orders. All taxa were included in the analyses: those which we sampled in phylogenies, but also those

for which sequences or phylogenies could not be obtained and which were included following the approaches described in the *colonisation and branching time* methods. A summary of diversity and sampling per archipelago is given in Extended Data Table 1.

Sampling completeness

In total, we produced new sequences from 252 new individuals, comprising 90 different species from 45 different genera, covering an additional 110 colonisation events that had never before been sampled (i.e. populations from islands where the species had not been sampled before). For at least 12 of these 90 species, we found no previous sequences on GenBank, including island endemics from Comoros, Galápagos, Rodrigues and São Tomé (Supplementary Table 5). The new sequences from 252 individuals increase the molecular sampling for extant colonisation events from 60% (223/373) to 89% (332/373). If we include historically extinct colonisations, we increased the molecular sampling from the existing 54% (269/502) of colonisation events to 75% (379/502). We also substantially increased molecular sampling of continental relatives, adding 78 new individuals from the continent or islands surrounding our archipelagos, covering 43 different species. The percentage of taxa sampled in phylogenies varied widely between archipelagos (Extended Data Table 1 and Fig. 1). For eight archipelagos (Bermuda, Fernando de Noronha, Pitcairn, Rapa Nui, Rodrigues, Saint Helena, Society Islands and Tonga) less than 50% of the species were sampled in phylogenies, and thus the majority of the species for these island groups were added with maximum ages and endemicity status. For 13 archipelagos, which accounted for more than a third of the total species, over 90% of the species were sampled in phylogenies.

DAISIE

We used the method DAISIE[10] (Dynamic Assembly of Islands through Speciation, Immigration, and Extinction) to estimate rates of species accumulation (colonisation, speciation and extinction) on the archipelagos. The model assumes that after the origin of an island, species can colonise from a mainland pool. Once a species has colonised, it may remain similar to its mainland ancestor (non-endemic species), become endemic through anagenetic speciation (new endemic species is formed without lineage splitting

on the island), split into new species via cladogenetic speciation and/or go extinct. A carrying capacity (i.e. maximum number of species each colonist lineage can attain) is implemented, such that rates of cladogenesis and colonisation decline with increasing number of species in the colonising clade.

The only effect of anagenesis under DAISIE is that the colonising species becomes endemic, because further anagenesis events on the endemic species do not leave a signature in the data. However, the rate of anagenesis is not systematically underestimated. Suppose the rate was higher; it would then follow that colonising species would also become endemic faster, and we would see more endemic species. Thus, the number of endemic species determines the rate of anagenesis, and DAISIE estimates the true rate of anagenesis without systematic bias. Further anagenesis events do not have an effect on the state variables, and hence do not enter the equations anymore.

In its parameterization of extinction, DAISIE accounts for the fact that there may have been several lineages that were present on the insular system in the past but which went completely extinct due to natural causes, leaving no extant descendants. Simulations have shown that the rate of natural extinction is usually well estimated in DAISIE (Methods section *Measuring precision and accuracy* and ref. [53]). Studies on phylogenies of single clades suggest that phylogenetic data on only extant species provide less information on extinction than on speciation (or rather diversification rates[54]). However, there is information-content in such data[55], especially when diversification dynamics are diversity-dependent[56]. Moreover, here we use colonisation times in addition to phylogenetic branching times to estimate extinction rates, and we are estimating hyperparameters that theory suggests correlate with extinction (i.e. area). Finally, we use data from many independent colonisations, which increases the power of our statistical method considerably, and decreases the bias, as ML is known to asymptotically provide unbiased estimates.

Estimating global hyperparameters

Our aim is to examine the dependencies of the parameters that govern species assembly (colonisation, extinction, cladogenesis, anagenesis (CES rates), and carrying capacity) on the features of archipelagos (area, isolation). We developed a new method to

estimate global hyperparameters that control the relationship between two key archipelago features (area and isolation) and archipelago-specific (local) CES rates. One can estimate directly from the global dataset the shape of the relationship between isolation and colonisation rate that maximizes the likelihood for the entire dataset.

Our method finds the hyperparameters that maximize the likelihood of the entire dataset, i.e. the sum of the log likelihoods for each archipelago. We tested the hypothesis that area and distance from the nearest mainland have an effect on CES rates (cladogenesis, anagenesis, extinction and colonisation). If an effect was identified we also estimated the scaling of the effect. We developed a set of *a priori* models where CES rates are affected by archipelago features as is often assumed in the island biogeography literature (Supplementary Table 1). For the *a priori* models, we considered that CES rates are determined by a power function of area or distance. In the power function, $par = par_0 I^h$, where par is the CES rate (e.g. local rate of colonisation), par_0 is the initial value of the biogeographical rate (e.g. global initial rate of colonisation), I is the physical variable (area or distance) and *h* is the strength of the relationship. The exponent *h* can be negative or positive depending on the nature of the relationship. par_0 and *h* are the hyperparameters. If the exponent *h* is estimated as zero, there is no relationship between I and the parameter. By including or excluding *h* from the different relationships we can compare different models with the effects switched on or off (Supplementary Table 1, e.g. in model M1 all relationships are estimated, but in model M2 the exponent of the relationship between anagenesis and distance is fixed to zero and thus anagenesis does not vary with distance).

In addition to the *a priori* models, we considered a set of *post hoc* models with alternative shapes of relationships. We fitted two types of *post hoc* models: power models and sigmoid models (Supplementary Table 1). In the *post hoc* power models we modelled all parameters as in the *a priori* models, except for cladogenesis: we allowed cladogenesis to be dependent both on area and distance. The reason for this is that we found that the predicted number of cladogenetic species under the *a priori* models were not as high as observed, so we examined whether including a positive effect of distance would improve the fit. We described the relationship between area, distance and cladogenesis using different functions – one model where there is an additive effect of area and distance (M15); and three models (M16, M17, M18) where the effect of area and distance is interactive. In addition, we fitted a model identical to M16 but with one

parameter less (M19). The reason for this was that this parameter (y) was being estimated as zero in M16

In the *post hoc* sigmoid models, we allowed the relationship between distance and a given parameter to follow a sigmoid rather a power function. The rationale for this was that we wanted to investigate whether for birds the effect of distance on a parameter only starts to operate after a certain distance from the mainland, as below certain geographical distances archipelagos are within easy reach for many bird species by flight so that at these distances the island behaves almost as part of the mainland from a bird's perspective. We fitted nine different sigmoid models (Supplementary Table 1), allowing cladogenesis, anagenesis and colonisation to vary with distance following a sigmoid function. The sigmoid function we used has an additional parameter in comparison to power functions.

In total we fitted 28 candidate models (14 *a priori*, 14 *post hoc*) to the global dataset using ML. We fitted each model using 20 initial sets of random starting parameters to reduce the risk of being trapped in local likelihood suboptima. We used the age of each archipelago (Extended Data Table 1) as the maximum age for colonisation. We assumed a global mainland species pool M of 1000 species. The product of M and the intrinsic rate of colonisation ($\sqrt{\gamma_0}$) is constant as long as M is large enough (larger than the number of island species), and thus the chosen value of M does not affect the results.

To decide which information criterion to use to select between different models we compared the performance of the BIC and the Akaike information criterion (AIC). We simulated 1,000 datasets each with models M9 and M19 and then fitted the M9, M14, M17 and M19 models to each of these datasets using two initial sets of starting parameters for each optimisation. We found that for datasets simulated using M9 an incorrect model was preferred using AIC in 10.4 % of cases, but only in 0.11 % of cases when using BIC. For datasets simulated using M19 an incorrect model was preferred 12.8 % of cases using AIC and 11.1 % of cases using BIC. We thus compared models using BIC, as this model has lower error rates.

An alternative approach to estimating hyperparameters would be to calculate CES rates and their uncertainty independently for each archipelago and to then conduct a meta-analysis of the resulting data, including archipelago area and isolation as predictors. However, errors in parameter estimates will vary, particularly because some

archipelagos have small sample sizes (only a few extant colonisation events, or none at all, e.g. Chagos) and are thus much less informative about underlying process[53]. Thus, maximizing the likelihood of all data sets together by estimating the hyperparameters (which is precisely our aim) is preferable. For completeness, we present CES rates estimated independently for each archipelago in Supplementary Table 6, excluding archipelagos with fewer than six species and for which we sampled less than 60% of the species in the phylogenies. However, as argued above we do not advocate using these parameter estimates for further analyses because the number of taxa for some of these archipelagos is still low and by excluding archipelagos with fewer than six taxa we cannot capture the lower part of the relationship between area/isolation and CES rates.

All DAISIE analyses were run using parallel computation on the high-performance computer clusters of the University of Groningen (Peregrine cluster) and the Museum für Naturkunde Berlin. The new version of the R package DAISIE is available on Github.

Randomization analysis

We conducted a randomization analysis to evaluate whether there is significant signal of a relationship between area and distance and local CES rates in our global dataset. We produced 1,000 datasets with the same phylogenetic data and archipelago ages as the global dataset, but randomly reshuffled archipelago area and D_m in each dataset. We then fitted the best *post hoc* model to each of these 1,000 randomized datasets. If the ML estimates of exponent hyperparameters (i.e. the strength of the relationship) in the randomized datasets were non-zero this would indicate that the method is finding evidence for a relationship even if there is none. If, on the other hand, non-zero hyperparameters are estimated in the real data but not in the randomized datasets, this would mean that there is information in the data regarding the putative relationships.

The randomization analysis showed that in global datasets with reshuffled areas and distances the exponent hyperparameters are estimated as zero in most cases, whereas in the empirical global dataset they are not (Extended Data Fig. 3).

A posteriori simulations

We simulated 1,000 phylogenetic global datasets (41 archipelagos each) with the ML

hyperparameters of the best *a priori* (M14) and *post hoc* models (M19). We first calculated the local CES rates for each archipelago based on their area and isolation and the hyperparameters for the model, and then used these CES rates as the parameters for the simulations using the DAISIE R package. The simulated data were used to measure bias and accuracy of the method, goodness of fit and the ability of our method to recover observed island biogeographic diversity patterns (see below).

Measuring precision and accuracy of method

DAISIE estimates CES rates with high precision and little bias[10,53]. We conducted parametric bootstrap analyses to assess whether the ability to estimate hyperparameters from global datasets is also good (Extended Data Fig. 2), and to obtain confidence intervals on parameter estimates (Extended Data Table 5). We used DAISIE to estimate hyperparameters from the M14 and M19 simulated datasets (1,000 replicates each). We measured precision and accuracy by comparing the distribution of parameters estimated from the 1,000 simulated data set with the real parameters used to simulate the same datasets. To check whether ML optimisations of the simulated global datasets converge to the same point in parameter space, we first performed a test on a subset of the simulated data. We ran optimisations with 10 random sets of initial starting values for each of 10 simulated datasets. All optimisations converged to the same likelihood and a very similar hyperparameter set; therefore, we are confident we found the global optimum for each simulated global dataset, even for models with many parameters.

Measuring goodness of fit

We measured how well the preferred models fitted the data using different approaches. First, we examined whether our models successfully reproduce the diversity patterns found on individual archipelagos. We calculated the total number of species, cladogenetic species and independent colonisations in each archipelago for each of the 1,000 simulated datasets. We then plotted these metrics versus the observed values in the empirical data (Extended Data Fig. 4 and Fig. 3). Our preferred models have a slight tendency to overpredict species richness when there are a few species and

underpredict it when there are many. We do not have a clear explanation for this. This slight deviation does not seem to be due to an additional area- or distance-dependence, so an explanation should be sought in other factors that we did not model. We note that the fact that all three plots show this tendency rather than just one is to be expected because the three metrics of species richness are not entirely independent, with total species richness being the sum of the other two.

Second, we examined whether the models successfully predict the empirical relationships between area, distance and diversity metrics (total species, cladogenetic species, and number of independent colonisations). We fitted generalised linear models (GLM) for each diversity metric, with quasipoisson family errors and log area (or distance) as predictors. We then repeated this across 1000 independent sets of simulated data for the 41 archipelagos and compared the mean of slopes and intercepts for archipelago area and archipelago isolation to the equivalent estimates for the empirical data (Fig. 4).

Third, we estimated the pseudo-R^2 of the best model (M19) as a measure of the model's explanatory power. We simulated two independent sets of 10,000 global datasets under M19 model (Set 1 and Set 2). We calculated the mean total number of species, number of cladogenetic species and colonisations for each archipelago across all datasets from Set 1. For each diversity metric we calculated a pseudo-R^2 (pseudo-R^2-observed) where the total sum of squares was obtained from the empirical data and the residual sum of squares was obtained as the difference between empirical values and expected values (i.e. the simulation means). As the model is inherently stochastic, even if the model is an accurate and complete reflection of the underlying processes then the pseudo-R^2 would tend to be < 1. To estimate the distribution of pseudo-R^2 expected under the model we treated the set 2 simulations as data and estimated the pseudo-R^2 for each (pseudo-R^2-simulated). We then calculated the ratio of the pseudo-R^2-observed values over the 10,000 pseudo-R^2-simulated values. A ratio approaching 1 would indicate that the model is explaining the observed data as well as the average dataset simulated under this process (Extended Data Fig. 5).

Sensitivity to alternative divergence times and tree topologies

Despite having sampled many new individuals from islands worldwide, given the wide

geographical scale of our study we still rely on sequence data for thousands of individuals submitted to GenBank over the years. Whenever multi-loci analyses including our focal taxa were available we used them, but these are rare (Extended Data Table 4). Therefore, the majority of our phylogenies are based on a small number of genes, and most on a single gene, cyt-*b*, which is the most widely sequenced mitochondrial marker in birds. Although some studies on island birds have shown that colonisation and diversification times derived from mitochondrial trees often do not differ much from those obtained using multiple loci (e.g. [57]), it is possible that for some cases the scaling and topologies of the trees might have been more accurate had we used multiple loci[58]. This is particularly relevant for recent island colonists, given incomplete lineage sorting[59]. An additional shortcoming of relying on published sequence data is that many of our DNA alignments often have substantial sections with missing data (e.g. because only one small section of the gene could be sequenced and was uploaded to GenBank), which has been shown to lead to biases in branch lengths and topology[60]. While future studies using phylogenomic approaches may address these issues, obtaining tissue samples for all these taxa will remain an obstacle for a long time.

Although DAISIE does not directly use topological information (only divergence times are used), it is possible that the true **topology** for a clade may differ from that of the gene tree we have estimated and this could have an impact on our results by a) affecting colonisation and branching times (addressed in the paragraph below); or b) by altering the number of colonisation events. Alternative topologies may have led to an increase or decrease in colonisation events – for instance, some species that appear to have colonised an archipelago only once may have colonised multiple times and if these re-colonisations are recent they may go undetected when using one or few loci. As with any phylogenetic study, we cannot rule out this possibility, but we assume that recent re-colonisation of the archipelagos in our dataset by the same taxon is rare, as these are all oceanic and isolated. For archipelago lineages with cladogenesis (26 out of 502 lineages), alternative topologies could include non-monophyly of island radiations, with the corollary being that they would be the result of multiple colonisation events. However, this seems improbable for these isolated and well-studied radiations, for which morphological evidence (e.g. HBW[37]) is consistent with their monophyly as supported by existing molecular data.

Regarding **scaling of divergence times**, we assessed how uncertainty in our estimated node ages could influence our results by running an analysis of 100 datasets. For each dataset we sampled the node ages (i.e. colonisation and branching times) at random from a uniform distribution centred on the posterior mean for that node in the BEAST tree and extending twice the length of the highest posterior density (HPD) interval. For example, for a node with a 95% HPD interval of 2-3 Ma in our trees, the uniform distribution was set to between 1.5 and 3.5 Ma. The HPD interval will capture uncertainty under the selected phylogenetic and substitution models for the loci we used, but we conduct our sensitivity analysis over a broader interval to accommodate the potential that the selected models and gene trees are inadequate. For cases where using this approach meant that the lower bound of the uniform distribution was lower than 0, we assigned a value of 0.00001 Ma to the lower bound. We fitted the nine best models to the 100 datasets using five initial starting parameters for each model (total 4,500 optimisations). We found that parameter estimates across the 100 datasets do not differ strongly from those in the main dataset (Supplementary Table 7). Importantly, model selection was unaffected, with the M19 model being selected for all 100 datasets. This is because a lot of the information used for model selection is coming from the other sources of information DAISIE uses (island age, number of species, endemicity status) rather than colonisation/branching times.

The ML parameters of the M19 model and the resulting area and isolation dependencies for datasets D1 to D6 - discussed below - are shown in Extended Data Fig. 6 and the DAISIE R objects including these alternative datasets are available in Mendeley Data (https://doi.org/10.17632/sy58zbv3s2.2).

To account for uncertainty in **rates of molecular evolution**, we repeated all BEAST dating analyses for markers that were not cyt-*b* using 1) the Weir and Schluter[50] cyt-*b* rate (dataset D1, equal to main dataset) and 2) marker-specific rates estimated by Lerner et al.[41], which are also widely used in the literature (dataset D2). Although the trees dated using the Lerner et al. rates provide younger ages, we found that the DAISIE results were very similar using either approach (same model preferred and similar parameters). Therefore, in the main text we only discuss the results of analyses of D1, i.e. applying Weir and Schluter's cyt-*b* rate to all markers.

For some taxa we did not use the stem age as the estimate of colonisation time, and instead used alternative nodes (see 'Colonisation and branching times' section). To

test whether our **choice of nodes** affects our main conclusions, we recoded all such taxa by extracting the stem ages and used these ages as an upper bound for colonisation (DAISIE MaxAge option). We fitted all 28 models to this new dataset (D3) and found that the M19 model is preferred and that the parameters and area/isolation relationships vary only slightly from those of the main analysis. We therefore conclude that our results are robust to the node selection approach.

If **extinction has been high on the mainland**, or if we failed to **sample the closest relatives** of the island taxa, this could lead to an overestimation of colonisation times when using the stem age as the precise time of colonisation. To investigate how this could have influenced our results, we ran analyses of datasets where we allow colonisation to have happened at any time since the stem age (i.e. the time of divergence from the nearest relative of the taxon on the mainland). For this we use the DAISIE options Endemic_MaxAge or NonEndemic_MaxAge, which integrate over all possible ages between the given maximum age and the present (or the first branching event within the archipelago for cases where cladogenesis has occurred). We repeated this analysis coding all stem ages as maximum ages (D4), or coding only the 25% older stem ages as maximum ages (to account for the fact that on older stems there is potential for there to be more bias) (D5). We also ran analyses on 100 datasets (D6) for which we assigned precise younger ages by randomly selecting a value between the stem age and the present (or crown age for cladogenetic groups). For all these datasets (D4-D6) we found that the same model (M19) is preferred, but the initial values of the biogeographical rates (cladogenesis, extinction, colonisation and anagenesis) are estimated to be higher than in the main dataset. Importantly, the exponent hyperparameters are similar to those in the main dataset, meaning that the shape of the relationships between parameters and area/isolation is not much affected (Extended Data Fig. 6). The only exception is perhaps anagenesis, for which the relationships vary more markedly – with isolated islands achieving very high rates for this parameter -, but still agreeing with our main conclusions. Anagenesis is in general the most difficult parameter to estimate[53]. Thus, our conclusions are robust to the colonisation times potentially being younger than those in our main dataset.

Sensitivity to archipelago selection and isolation metrics

The results of the following sensitivity analyses are presented in Supplementary Data 3 and the DAISIE R objects including these alternative datasets are available in Mendeley Data (https://doi.org/10.17632/sy58zbv3s2.2).

To test whether the **inclusion of both true archipelagos and single islands** in our dataset could affect the results, we repeated analyses excluding single island units and found that the same model is preferred. The estimated initial rate of cladogenesis (λ^c_0) is higher if we exclude single islands, but this parameter is not different from a distribution of parameters estimated from datasets generated using a stratified-random sampling of both archipelagos and single islands.

Alternative isolation metrics to D_m have been shown to explain varying and often higher amounts of variation in species richness on islands[61]. We tested two alternative metrics: distance to the nearest larger or equivalent-sized landmass (D_b), and the mean between D_m and D_b (metrics given in Supplementary Data 2). We found that the same DAISIE model with very similar parameters was preferred in both cases, and thus we used only the D_m metric, as this is more similar to the original model of MacArthur & Wilson.

The **Mascarenes** (Mauritius Isl., Reunion and Rodrigues) are often treated as a single biogeographical unit in analyses. We chose to analyse them as independent units because a) the distance between islands is much greater than our threshold for archipelago definition (more than 500 km between Mauritius Isl. and Rodrigues; more than 170 km between Reunion and Mauritius Isl.); b) only two species of our target group are shared between the islands (*Terpsiphone bourbonnensis* found in Mauritius Isl. and Reunion; and *Psittacula eques* found in Mauritius Isl. and extirpated from Reunion), suggesting low connectivity; c) while there are three clades whose branching events took place within the Mascarenes (*Coracina*, *Pezophaps/Raphus* and *Zosterops*), the remaining species result from independent colonisations suggesting that the three islands behave mostly as three different biogeographical units. We nevertheless ran an analysis treating the islands as a single archipelagic unit and found that the same model was preferred and with similar parameter estimates, and we therefore discuss only the results treating them as separate.

31. Plummer, P. S. & Belle, E. R. Mesozoic tectono-stratigraphic evolution of the Seychelles microcontinent. *Sediment. Geol.* **96**, 73–91 (1995).
32. Weigelt, P., Jetz, W. & Kreft, H. Bioclimatic and physical characterization of the world's islands.

Proc. Natl. Acad. Sci. U. S. A. **110**, 15307–15312 (2013).
33. Norder, S. J. *et al.* Beyond the Last Glacial Maximum: Island endemism is best explained by long-lasting archipelago configurations. *Glob. Ecol. Biogeogr.* **28**, 184-197(2018)
34. Thomson, J. & Walton, A. Redetermination of chronology of Aldabra atoll by 230Th/ 234U dating. *Nature* **240**, 145–146 (1972).
35. Price, J. P. & Clague, D. A. How old is the Hawaiian biota? Geology and phylogeny suggest recent divergence. *Proc. Biol. Sci.* **269**, 2429–35 (2002).
36. Valente, L. *et al.* Equilibrium bird species diversity in Atlantic islands. *Curr. Biol.* **27**, 1660–1666 (2017).
37. *Handbook of the Birds of the World Alive*. (Lynx Edicions, 2018).
38. Valente, L., Etienne, R. S. & Dávalos, L. M. Recent extinctions disturb path to equilibrium diversity in Caribbean bats. *Nat. Ecol. Evol.* **1**, 26 (2017).
39. Steadman, D. W. *Extinction and biogeography of tropical Pacific birds*. (University of Chicago Press, 2006).
40. Cheke, A. & Hume, J. P. *Lost land of the Dodo: The ecological history of Mauritius, Réunion and Rodrigues*. (Bloomsbury Publishing, 2010).
41. Lerner, H. R. L., Meyer, M., James, H. F., Hofreiter, M. & Fleischer, R. C. Multilocus resolution of phylogeny and timescale in the extant adaptive radiation of Hawaiian honeycreepers. *Curr. Biol.* **21**, 1838–44 (2011).
42. Rando, J. C., Pieper, H., Olson, S. L., Pereira, F. & Alcover, J. A. A new extinct species of large bullfinch (Aves: Fringillidae: *Pyrrhula*) from Graciosa Island (Azores, North Atlantic Ocean). *Zootaxa* **4282**, 567–583 (2017).
43. Illera, J. C., Rando, J. C., Richardson, D. S. & Emerson, B. C. Age, origins and extinctions of the avifauna of Macaronesia: a synthesis of phylogenetic and fossil information. *Quat. Sci. Rev.* **50**, 14–22 (2012).
44. Hume, J. P., Martill, D. & Hing, R. A terrestrial vertebrate palaeontological review of Aldabra Atoll, Aldabra Group, Seychelles. *PLoS One* **13**, e0192675 (2018).
45. Cheke, A. S. Extinct birds of the Mascarenes and Seychelles–a review of the causes of extinction in the light of an important new publication on extinct birds. *Phelsuma* **21**, 4–19 (2013).
46. Hume, J. P. & Walters, M. *Extinct birds*. (A&C Black, 2012).
47. Kearse, M. *et al.* Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647–9 (2012).
48. Bouckaert, R. *et al.* BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* **10**, e1003537 (2014).
49. Posada, D. jModelTest: Phylogenetic Model Averaging. *Mol Biol Evol* **25**, 1253–1256 (2008).
50. Weir, J. T. & Schluter, D. Calibrating the avian molecular clock. *Mol. Ecol.* **17**, 2321–2328 (2008).
51. Field, D. J. *et al.* Timing the extant avian radiation: the rise of modern birds, and the importance of modeling molecular rate variation. *PeerJ Prepr.* **7**, e27521v1 (2019).
52. Cicero, C. & Johnson, N. K. Higher-level phylogeny of New World vireos (Aves: Vireonidae) based on sequences of multiple mitochondrial DNA genes. *Mol. Phylogenet. Evol.* **20**, 27–40 (2001).
53. Valente, L., Phillimore, A. & Etienne, R. S. Using molecular phylogenies in island biogeography: it's about time. *Ecography (Cop.).* **41**, 1684–1686 (2018).
54. Rabosky, D. L. Extinction rates should not be estimated from molecular phylogenies. *Evolution.* **64**, 1816–1824 (2010).
55. Nee, S., May, R. M. & Harvey, P. H. The reconstructed evolutionary process. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **344**, 305–11 (1994).
56. Etienne, R. S. *et al.* Diversity-dependence brings molecular phylogenies closer to agreement with the fossil record. *Proc. R. Soc. B Biol. Sci.* **279**, 1300–1309 (2012).
57. Stervander, M. *et al.* Disentangling the complex evolutionary history of the Western Palearctic blue tits (*Cyanistes* spp.) - phylogenomic analyses suggest radiation by multiple colonisation events and subsequent isolation. *Mol. Ecol.* **24**, 2477–2494 (2015).
58. Ogilvie, H. A., Heled, J., Xie, D. & Drummond, A. J. Computational performance and statistical accuracy of *BEAST and comparisons with other methods. *Syst. Biol.* **65**, 381–396 (2016).
59. Maddison, W. P. & Knowles, L. L. Inferring phylogeny despite incomplete lineage sorting. *Syst. Biol.* **55**, 21–30 (2006).
60. Lemmon, A. R., Brown, J. M., Stanger-Hall, K. & Lemmon, E. M. The effect of ambiguous data on phylogenetic estimates obtained by maximum likelihood and Bayesian inference. *Syst. Biol.* **58**, 130–145 (2009).
61. Weigelt, P. & Kreft, H. Quantifying island isolation–insights from global patterns of insular plant

species richness. *Ecography (Cop.).* **36**, 417–429 (2013).
62. Nielson, D. L. & Sibbett, B. S. Geology of ascension Island, South Atlantic Ocean. *Geothermics* **25**, 427–448 (1996).
63. Ramalho, R. S. *et al.* Emergence and evolution of Santa Maria Island (Azores)—The conundrum of uplifted islands revisited. *Geol. Soc. Am. Bull.* **129**, 372–390 (2017).
64. Hearty, P. J. & Olson, S. L. Geochronology, biostratigraphy, and changing shell morphology in the land snail subgenus Poecilozonites during the Quaternary of Bermuda. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **293**, 9–29 (2010).
65. Carracedo, J. C. & Troll, V. R. *The Geology of the Canary Islands*. (Elsevier, 2016).
66. Ramalho, R. *Building the Cape Verde Islands*. (Springer, 2011).
67. Eisenhauer, A., Heiss, G. A., Sheppard, C. R. C. & Dullo, W. C. Reef and island formation and Late Holocene sea-level changes in the Chagos islands. *Ecol. Chagos Archipel.* 21–31 (1999).
68. Campbell, H. J. Fauna and flora of the Chatham Islands: less than 4 my old. *Geol. genes* **97**, 15–16 (1998).
69. Bullough, F. History and Geology of Christmas Island. *Geological Society of London blog* https://blog.geolsoc.org.uk/2013/12/18/door-18-his (2013). Available at: https://blog.geolsoc.org.uk/2013/12/18/door-18-history-and-geology-of-christmas-island/.
70. Castillo, P. *et al.* Anomalously young volcanoes on old hot-spot traces: I. Geology and petrology of Cocos Island. *Geol. Soc. Am. Bull.* **100**, 1400–1414 (1988).
71. Woodroffe, C. D., Veeh, H. H., Falkland, A. C., McLean, R. F. & Wallensky, E. Last interglacial reef and subsidence of the Cocos (Keeling) Islands, Indian Ocean. *Mar. Geol.* **96**, 137–143 (1991).
72. Nougier, J., Cantagrel, J. M. & Karche, J. P. The Comores archipelago in the western Indian Ocean: volcanology, geochronology and geodynamic setting. *J. African Earth Sci.* **5**, 135–144 (1986).
73. Almeida, F. The Fernando de Noronha archipelago. in *Sítios Geológicos e Paleontológicos do Brasil* (ed. Schobbenhaus,C.; Campos,D.A.; Queiroz,E.T.; Winge,M.; Berbert-Born, M.) http://www.unb.br/ig/sigep/sitio066/sitio066englis (2000).
74. Ali, J. R. & Aitchison, J. C. Exploring the combined role of eustasy and oceanic island thermal subsidence in shaping biodiversity on the Galápagos. *J. Biogeogr.* **41**, 1227–1241 (2014).
75. Ryan, P. G. Tristan da Cunha and Gough Island. in *Encyclopedia of Islands* (eds. Gillespie, R. & Clague, D.) 929–932 (University of California Press, 2009).
76. Batiza, R. Petrology and chemistry of Guadalupe Island: An alkalic seamount on a fossil ridge crest. *Geology* **5**, 760–764 (1977).
77. Stuessy, T. F., Foland, K. A., Sutter, J. F., Sanders, R. W. & O., M. S. Botanical and geological significance of potassium-argon dates from the Juan Fernandez islands. *Science.* **225**, 49–51 (1984).
78. McDougall, I., Embleton, B. J. J. & Stone, D. B. Origin and evolution of Lord Howe Island, Southwest Pacific Ocean. *J. Geol. Soc. Aust.* **28**, 155–176 (1981).
79. Mata, J. *et al.* O arquipélago da Madeira - Geografia de Portugal. in *Geologia de Portugal* (eds. Dias, R., Araujo, A., Terrinha, P. & Kullberg, J.) **2**, 691–746 (Escolar Editora, 2013).
80. Guille, G. *et al.* Les marquises (Polynésie Françaises): un archipel intraocéanique atypique. *Géologie la Fr.* **2**, 5–36 (2002).
81. Montaggioni, L. & Nativel, P. *La Reunion, Ile Maurice. Géologie et aperçus biologiques, plantes et animaux*. (Masson, 1988).
82. Grandcolas, P. *et al.* New Caledonia: a very old Darwinian island? *Philos. Trans. R. Soc. B Biol. Sci.* **363**, 3309–3317 (2008).
83. Anthoni, J. Geography and geology of Niue. http://www.seafriends.org.nz/niue/geo.htm (2005). Available at: http://www.seafriends.org.nz/niue/geo.htm.
84. Jones, J. G. & McDougall, I. Geological history of Norfolk and Philip islands, southwest Pacific ocean. *J. Geol. Soc. Aust.* **20**, 240–254 (1973).
85. Suzuki, M., Taisuke, S. & Hideo, T. *Nomination of the Ogasawara Islands for inscription on the World Heritage List*. (Government of Japan, 2010).
86. Neall, V. E. & Trewick, S. A. The age and origin of the Pacific islands: a geological overview. *Philos. Trans. R. Soc. B Biol. Sci.* **363**, 3293–3308 (2008).
87. Hekinian, R. *et al.* The Pitcairn hotspot in the South Pacific: Distribution and composition of submarine volcanic sequences. *J. Volcanol. Geotherm. Res.* **121**, 219–245 (2003).
88. Vezzoli, L. & Acocella, V. Easter Island, SE Pacific: An end-member type of hotspot volcanism. *Bull. Geol. Soc. Am.* **121**, 869–886 (2009).
89. Gillot, P.-Y., Lefèvre, J.-C. & Nativel, P.-E. Model for the structural evolution of the volcanoes of Réunion Island. *Earth Planet. Sci. Lett.* **122**, 291–302 (1994).

90. Safford, R. & Hawkins, F. *The birds of Africa: Volume VIII: The Malagasy Region: Madagascar, Seychelles, Comoros, Mascarenes*. **8**, (A&C Black, 2013).
91. Baker, I., Gale, N. H. & Simons, J. Geochronology of the St Helena volcanoes. *Nature* **215**, 1451–1456 (1967).
92. Duncan, R. A. Radiometric ages from volcanic rocks along the New Hebrides-Samoa lineament. in *Investigations of the Northern Melanesian Borderland, Earth Science Series* (Circum Pacific Council Publications, 1985).
93. Lee, D. C., Halliday, A. N., Fitton, J. G. & Poli, G. Isotopic variations with distance and time in the volcanic islands of the Cameroon line: evidence for a mantle plume origin. *Earth Planet. Sci. Lett.* **123**, 119–138 (1994).
94. Geldmacher, J., Hoernle, K., Van Den Bogaard, P., Zankl, G. & Garbe-Schönberg, D. Earlier history of the ≥70-Ma-old Canary hotspot based on the temporal and geochemical evolution of the Selvagen Archipelago and neighboring seamonts in the Eastern North Atlantic. *J. Volcanol. Geotherm. Res.* **111**, 55–87 (2001).
95. Clouard, V. & Bonneville, A. Ages of seamounts, islands, and plateaus on the Pacific plate. in *Plates, plumes and paradigms* (eds. Foulger, G. R., Natland, J. H., Presnall, D. C. & Anderson, D. L.) 71–90 (2005). doi:10.1130/0-8137-2388-4.71
96. Bohrson, W. a. *et al.* Prolonged history of silicic peralkaline volcanism in the eastern Pacific Ocean. *J. Geophys. Res.* **101**, 11457 (1996).
97. Kroenke, L. W. Plate tectonic development of the western and southwestern Pacific: mesozoic to the present. in *The origin and evolution of Pacific Island biotas, New Guinea to Eastern Polynesia: patterns and processes* (eds. Keast, A. & Miller, S.) 19–34 (1996).
98. Ollier, C. D. Geomorphology of South Atlantic volcanic islands. Part I: the Tristan da Cunha group. *Zeitschrift fur Geomorphol.* **28**, 367–382 (1984).
99. Kocher, T. D. *et al.* Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. *Proc. Natl. Acad. Sci. U. S. A.* **86**, 6196–6200 (1989).
100. Dietzen, C., Witt, H.-H. & Wink, M. The phylogeographic differentiation of the European robin *Erithacus rubecula* on the Canary Islands revealed by mitochondrial DNA sequence data and morphometrics: evidence for a new robin taxon on Gran Canaria? *Avian Sci.* **3**, 115–132 (2003).
101. Edwards, S. V, Arctander, P. & Wilson, A. C. Mitochondrial resolution of a deep branch in the genealogical tree for perching birds. *Proc. Biol. Sci.* **243**, 99–107 (1991).
102. Helm-Bychowski, K. & Cracraft, J. Recovering phylogenetic signal from DNA sequences: relationships within the corvine assemblage (class Aves) as inferred from complete sequences of the mitochondrial DNA cytochrome-b gene. *Mol. Biol. Evol.* **10**, 1196–1214 (1993).
103. Warren, B. H., Bermingham, E., Bowie, R. C. K., Prys-Jones, R. P. & Thébaud, C. Molecular phylogeography reveals island colonization history and diversification of western Indian Ocean sunbirds (*Nectarinia*: Nectariniidae). *Mol. Phylogenet. Evol.* **29**, 67–85 (2003).
104. Farrington, H. L., Lawson, L. P., Clark, C. M. & Petren, K. The evolutionary history of Darwin's finches: speciation, gene flow, and introgression in a fragmented landscape. *Evolution* **68**, 2932-2944 (2014).
105. Warren, B. H. *et al.* Hybridization and barriers to gene flow in an island bird radiation. *Evolution.* **66**, 1490–1505 (2012).
106. Warren, B. H., Bermingham, E., Prys-Jones, R. P. & Thebaud, C. Tracking island colonization history and phenotypic shifts in Indian Ocean bulbuls (*Hypsipetes*: Pycnonotidae). *Biol. J. Linn. Soc.* **85**, 271–287 (2005).
107. Andersen, M. J., Hosner, P. A., Filardi, C. E. & Moyle, R. G. Phylogeny of the monarch flycatchers reveals extensive paraphyly and novel relationships within a major Australo-Pacific radiation. *Mol. Phylogenet. Evol.* **83**, 118–136 (2015).
108. Sari, E. H. R. & Parker, P. G. Understanding the colonization history of the Galápagos flycatcher (*Myiarchus magnirostris*). *Mol. Phylogenet. Evol.* **63**, 244–54 (2012).
109. Chaves, J. A., Parker, P. G. & Smith, T. B. Origin and population history of a recent colonizer, the yellow warbler in Galápagos and Cocos Islands. *J. Evol. Biol.* **25**, 509–21 (2012).
110. Martínez-Gómez, J. E., Barber, B. R. & Peterson, a T. Phylogenetic position and generic placement of the Socorro wren (*Thryomanes sissonii*). *Au* **122**, 50–56 (2005).
111. Warren, B. H., Bermingham, E., Prys-Jones, R. P. & Thébaud, C. Immigration, species radiation and extinction in a highly diverse songbird lineage: White-eyes on Indian Ocean islands. *Mol. Ecol.* **15**, 3769–3786 (2006).
112. McGuire, J. A. *et al.* Molecular phylogenetics and the diversification of hummingbirds. *Curr. Biol.* **24**, 910–916 (2014).

113. Derryberry, E. P. *et al.* Lineage diversification and morphological evolution in a large-scale continental radiation: The neotropical ovenbirds and woodcreepers (Aves: Furnariidae). *Evolution.* **65**, 2973–2986 (2011).
114. Jønsson, K. A. *et al.* A supermatrix phylogeny of corvoid passerine birds (Aves: Corvides). *Mol. Phylogenet. Evol.* **94**, 87–94 (2016).
115. Scofield, R. P. *et al.* The origin and phylogenetic relationships of the New Zealand ravens. *Mol. Phylogenet. Evol.* **106**, 136–143 (2017).
116. Cibois, A., Thibault, J. C., Bonillo, C., Filardi, C. E. & Pasquet, E. Phylogeny and biogeography of the imperial pigeons (Aves: Columbidae) in the Pacific Ocean. *Mol. Phylogenet. Evol.* **110**, 19–26 (2017).
117. Friis, G., Aleixandre, P., Rodríguez-Estrella, R., Navarro-Sigüenza, A. G. & Milá, B. Rapid postglacial diversification and long-term stasis within the songbird genus *Junco*: phylogeographic and phylogenomic evidence. *Mol. Ecol.* **25**, 6175–6195 (2016).
118. Marki, P. Z. *et al.* Supermatrix phylogeny and biogeography of the Australasian Meliphagides radiation (Aves: Passeriformes). *Mol. Phylogenet. Evol.* **107**, 516–529 (2017).
119. Fuchs, J. *et al.* Long-distance dispersal and inter-island colonization across the western Malagasy Region explain diversification in brush-warblers (Passeriformes: *Nesillas*). *Biol. J. Linn. Soc.* **119**, 873–889 (2016).
120. Cibois, A. *et al.* Phylogeny and biogeography of the fruit doves (Aves: Columbidae). *Mol. Phylogenet. Evol.* **70**, 442–453 (2014).
121. Carmi, O., Witt, C. C., Jaramillo, A. & Dumbacher, J. P. Phylogeography of the Vermilion Flycatcher species complex: Multiple speciation events, shifts in migratory behavior, and an apparent extinction of a Galápagos-endemic bird species. *Mol. Phylogenet. Evol.* **102**, 152–173 (2016).
122. Cornetti, L. *et al.* The genome of the 'great speciator' provides insights into bird diversification. *Genome Biol. Evol.* **7**, 2680–2691 (2015).

Data availability

New sequence data produced for this study have been deposited in GenBank with the accession codes: MH307408- MH307656. The following datasets have been deposited in Mendeley: DNA alignments (https://doi.org/10.17632/vf95364vx6.1), new phylogenetic trees produced for this study (https://doi.org/10.17632/p6hm5w8s3b.2) and DAISIE R objects (https://doi.org/10.17632/sy58zbv3s2.2). The 11 previously published trees are available upon request.

Code availability

The custom computer code used for this study is freely available in the DAISIE R package (https://github.com/rsetienne/DAISIE).

Acknowledgments

We thank the skilled guides and field assistants who helped with sample collection in the field; and the ornithologists and collection curators who were kind enough to reply to requests for material. For support/advice: Thomas von Rintelen, Kristina von Rintelen, Christine Zorn. For comments on the manuscript: Alex Pigot. For providing samples or DNA sequences: Nancy Bunbury (Seychelles Islands Foundation) who

organized sample loans of Aldabra island; Janske van de Crommenacke, Jim Groombridge, Hazel Jackson. For sharing data on extinct species: Josep Antoni Alcover, Juan Carlos Rando, Ferran Sayol, Søren Faurby. For permission to use photographs or illustrations: Cláudia Baeta, Martijn Hammers, Julian Hume, Dubi Shapiro, Juan Varela, Pedro Cascão. For expertise on island geological ages: Paul Hearty, Robert Stern, Mark Reagan. For providing phylogenetic data: Alice Cibois, Jimmy McGuire, Heather Lerner, Petter Marki, Borja Milá, Guillermo Friis, Jérôme Fuchs, John P. Dumbacher, Ore Carmi. For map data: Patrick Weigelt. For permission to obtain new samples or to access existing samples, and for logistic support: São Tomé e Príncipe: Arlindo Carvalho and the Department of the Environment; Equatorial Guinea: José Obiang, Noélia Calvo, the Universidad Nacional de Guinea Ecuatorial for Bioko and Annobón samples; Seychelles: the Ministry of Environment, Energy and Climate Change of the Republic of Seychelles, the Seychelles Bureau of Standards, BirdLife Seychelles, Seychelles Islands Foundation; Comoros: Centre National de Documentation et de Recherche Scientifique (Grande Comore & Anjouan), Action Comores, Direction de l'Agriculture et de la Foret (Mayotte); Madagascar: Ministere des Eaux et Forets (Madagascar), the Madagascar Institute pour la Conservation des Ecosystemes Tropicaux; Mauritius: Mauritius National Parks and Conservation Service, Mauritius Wildlife Foundation; New Caledonia: Olivier Hébert, Waifite Waheoneme, Nicholas Clark, the Direction de L'Environment (South Province), Direction du Développement Economique (Loyalty Islands Province), local chiefs and landowners; Morocco: Moroccan Environment Ministry; Cape Verde: Cape Verde Agriculture and Environment Ministry; Cameroon: Francis Njie and the Limbe Botanical and Zoological Garden; Gabon: Station de Recherche de l'IRET at Ipassa-Makokou; Angola: Fernanda Lages (ISCED-Huíla); Spain: the regional governments of Andalucía and the Canary Islands; Portugal: regional governments of Madeira and the Azores. Museum samples: Department of Ornithology and Mammalogy of the California Academy of Sciences (Laura Wilkinson & Maureen Flannery) for loaning Galápagos samples; Natural History Museum at Tring (Mark Adams) for loaning Comoros samples; the Stuttgart State Museum of Natural History for loaning stonechat samples from Madagascar. Sebastian Block assisted with cluster analyses at the Museum für Naturkunde. The Center for Information Technology of the University of Groningen provided support and access to the Peregrine high-performance computing cluster.

L.V. was funded by the German Science Foundation (DFG Research grant VA 1102/1-1), the Alexander von Humboldt Foundation, the Brandenburg Postdoc Prize 2015 and by a VIDI grant from the Netherlands Organisation for Scientific Research (NWO); R.S.E. by a NWO VICI grant; M.M. by the Portuguese Science and Technology Foundation (Post-doctoral grant: SFRH/BPD/100614/2014); S.M.C by the National Geographic Society (CRE grant # 9383-13); J.C.I. by the Spanish Ministry of Science, Innovation and Universities (Ref.: PGC2018-097575-B-I00) and by a GRUPIN research grant from the Regional Government of Asturias (Ref.: IDI/2018/000151); C.T. by the 'Laboratoire d'Excellence' TULIP (ANR-10-LABX-41).

Author contributions

L.V., A.B.P. and R.S.E. designed the study, developed the analytical framework and performed statistical analyses. L.V. compiled the data, conducted most of the analyses and wrote the first draft. R.S.E. developed the likelihood method. A.B.P. and R.S.E. contributed substantially to the writing. M.M., B.H.W., S.M.C., J.C.I. and C.T. provided expertise on islands birds and collected bird tissue samples, as well as molecular and/or phylogenetic data. K.H. and J.C.I. performed laboratory work. R.T. contributed to molecular analyses. T.A. performed analyses. All authors commented on the draft.

Supplementary information is available for this paper.

Competing interests The authors declare no competing interests.

Correspondence and requests for materials should be addressed to L.V.

Extended Data Figure Legends

Extended Data Figure 1 | Variation of cladogenesis with isolation and area. Contour plot showing how the local rate of cladogenesis varies with area and distance from the nearest mainland (D_m) assuming the ML global hyperparameters of the M19 model (equations describing the relationships given in Supplementary Table 1). Numbers correspond to the archipelago numbers from Fig. 1, and show the local cladogenesis rates for each of the archipelagos in our dataset. Area in log scale.

Extended Data Figure 2 | Bootstrap precision estimates of the parameters of the M19 model. Parametric bootstrap analysis fitting the M19 model to 1,000 global data sets simulated with ML parameters of the M19 model. Plots are frequency histograms of estimated parameters. Black lines show the median estimated values across all simulations and the blue lines the simulated values. Dashed lines show 2.5 – 97.5 percentiles. Parameters explained in Supplementary Table 1. Bootstrap parameter estimates for the M14 model are shown in Extended Data Table 5.

Extended Data Figure 3 | Randomization analysis of the M19 model. Distribution of global hyperparameters estimated from each of 1,000 datasets with the same phylogenetic data as our main global dataset but randomly reshuffling archipelago area and isolation among the 41 archipelagos in the data. Grey histograms show DAISIE ML parameter estimates for the M19 model. Red arrow shows the estimated parameter from the real data. In the majority of cases, the hyperparameters describing the exponent of the power models (x, α, β and d_0) are estimated as zero in the reshuffled datasets, which is not the case in the real data (red). Parameters explained in Supplementary Table 1.

Extended Data Figure 4 | Goodness of fit of the preferred model (M19). Plots show observed total number of species, cladogenetic species and colonisations versus those simulated under the model. Median and 95% percentiles shown for 1000 simulations of each archipelago. Selected archipelagos are highlighted in colour. Dashed line is y=x. See also Fig. 3.

Extended Data Figure 5 | Ratio of pseudo-R^2-observed over pseudo-R^2-simulated. Based on 10,000 datasets simulated under M19 model. A ratio centred on 1 would indicate that the model explains the observed data as well as it is able to explain the average dataset simulated under the ML parameters.

Extended Data Figure 6 | Sensitivity to colonisation and branching times. a, ML parameter estimates of the M19 model (preferred model) for datasets differing in colonisation and branching times. D6 represents 100 datasets, therefore, the 2.5 and 97.5 percentiles are shown. Parameter symbols as in Supplementary Table 1. **b,** Estimated relationships between island area and isolation and local island biogeography parameters for each dataset. Under the M19 model, cladogenesis rate increases with both area and isolation, and thus plots for more (far, 5,000 km) and less (near, 50 km) isolated islands are shown.

Extended Data table titles and footnotes

Extended Data Table 1 | Archipelago characteristics and references for the island geological ages. More data in Supplementary Data 2. For archipelagos closer to Madagascar, New Guinea or New Zealand than to the continent, we use those islands as the mainland.

Footnote:

* [34] proposed an age of 0.125 Ma, but we used older age, see Methods.

† At least 2 Ma; Paul Hearty pers. comm.

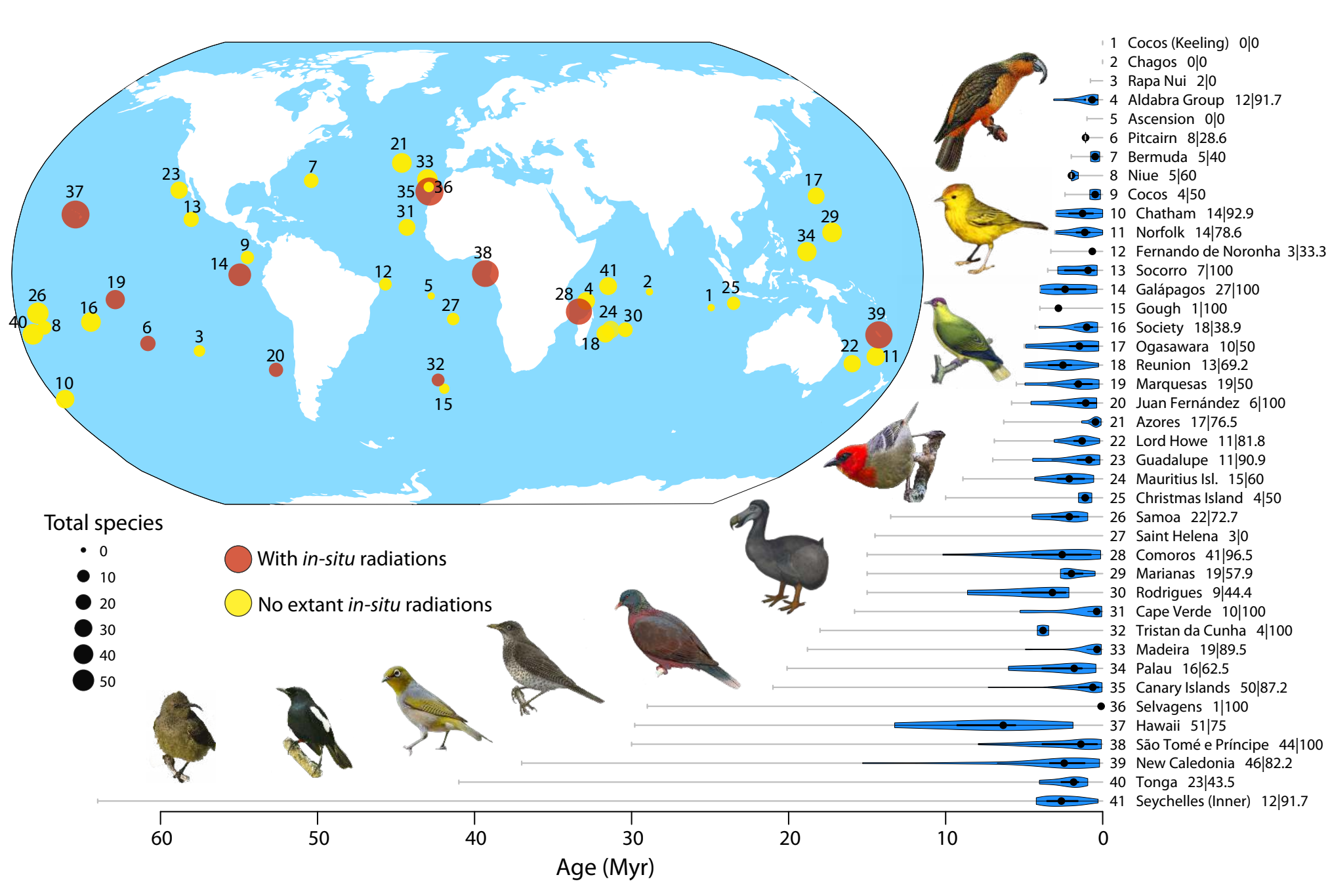
‡ Robert Stern & Mark K. Reagan pers. comm.

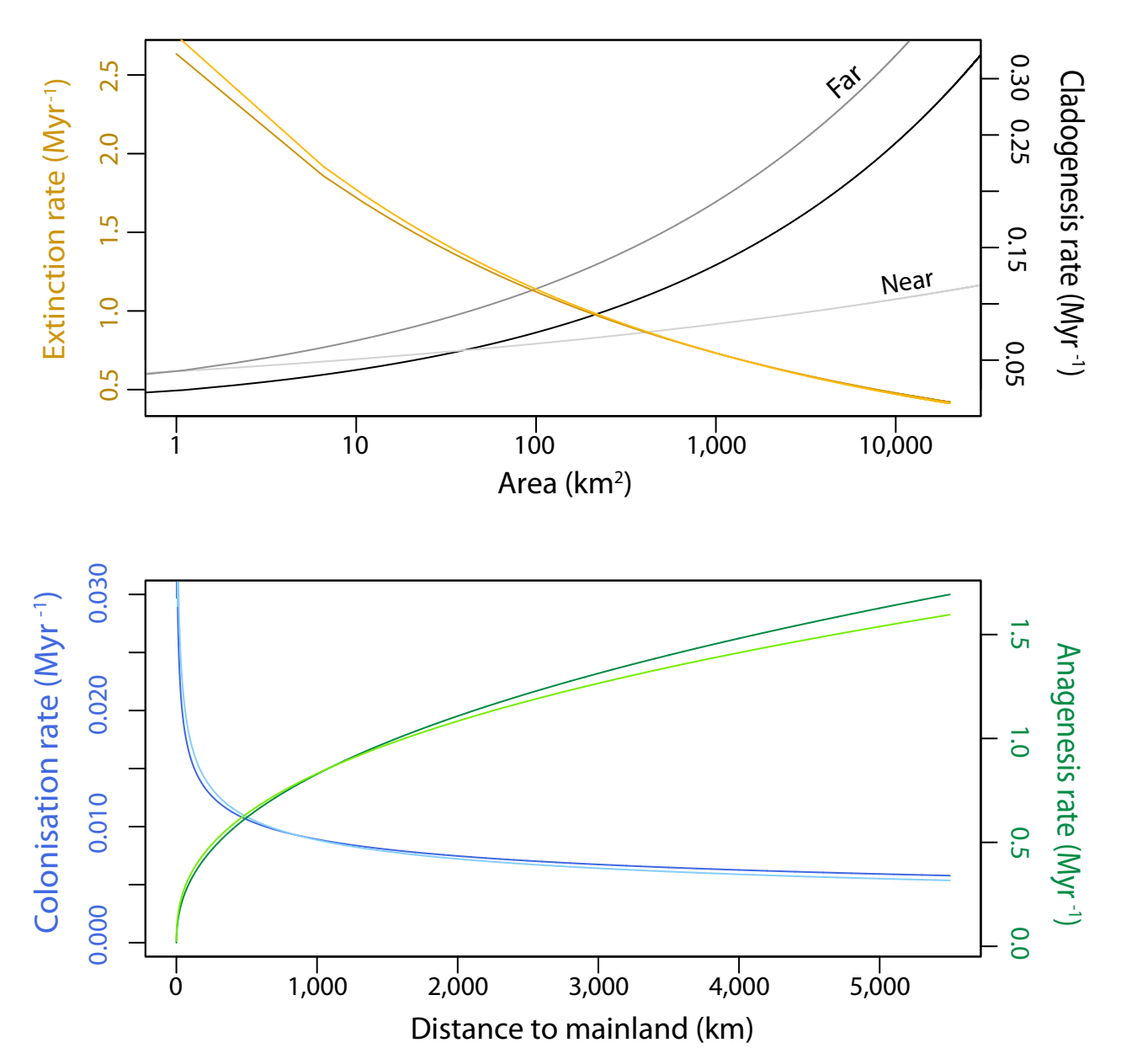
Extended Data Table 2 | Primer sequences used in this study.

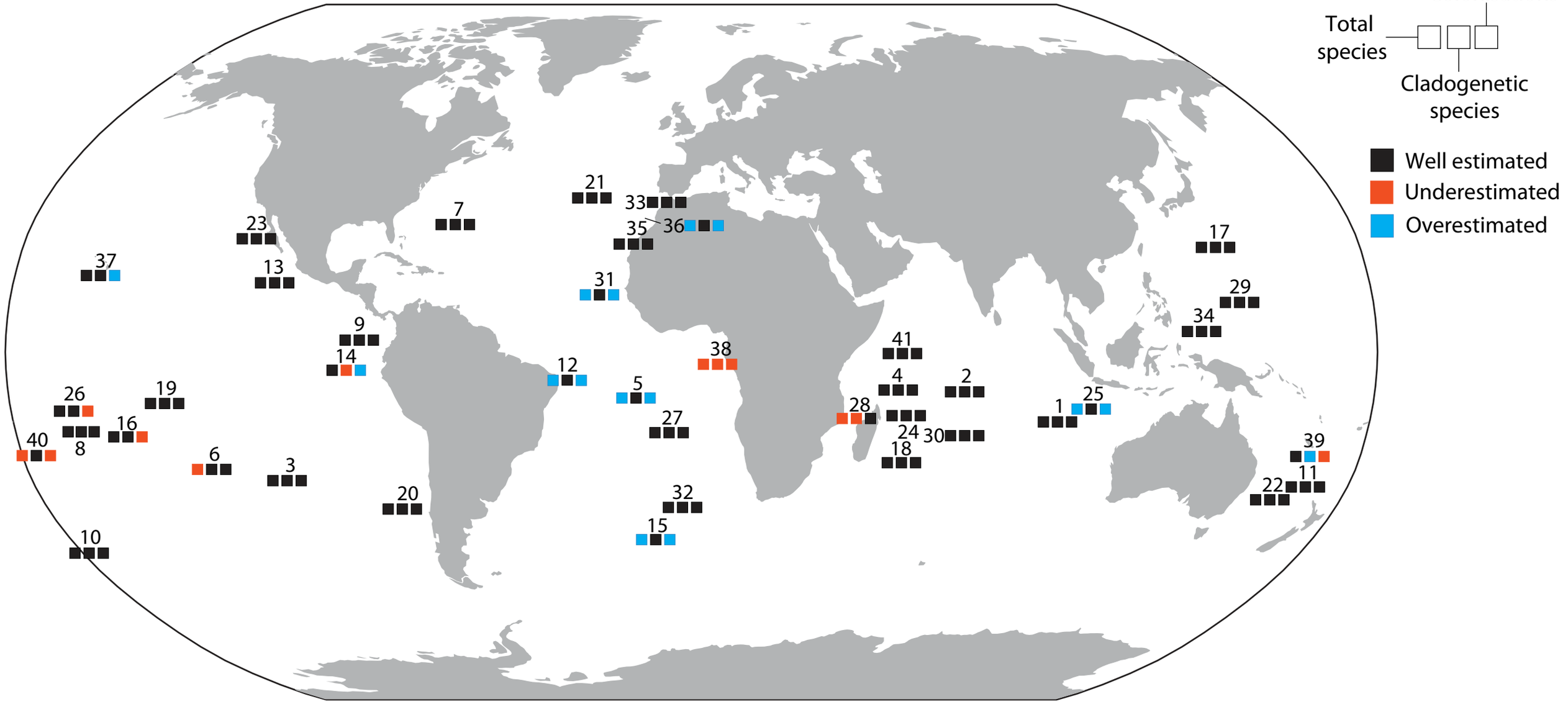
Extended Data Table 3 | The 80 alignments used in the new phylogenetic analyses. Main source of sequences is GenBank or the new sequences produced for this study, except for the cases noted in the table, where a matrix was directly obtained from a specific study. Details on molecular rates and molecular models applied to each alignment in Supplementary Table 4.

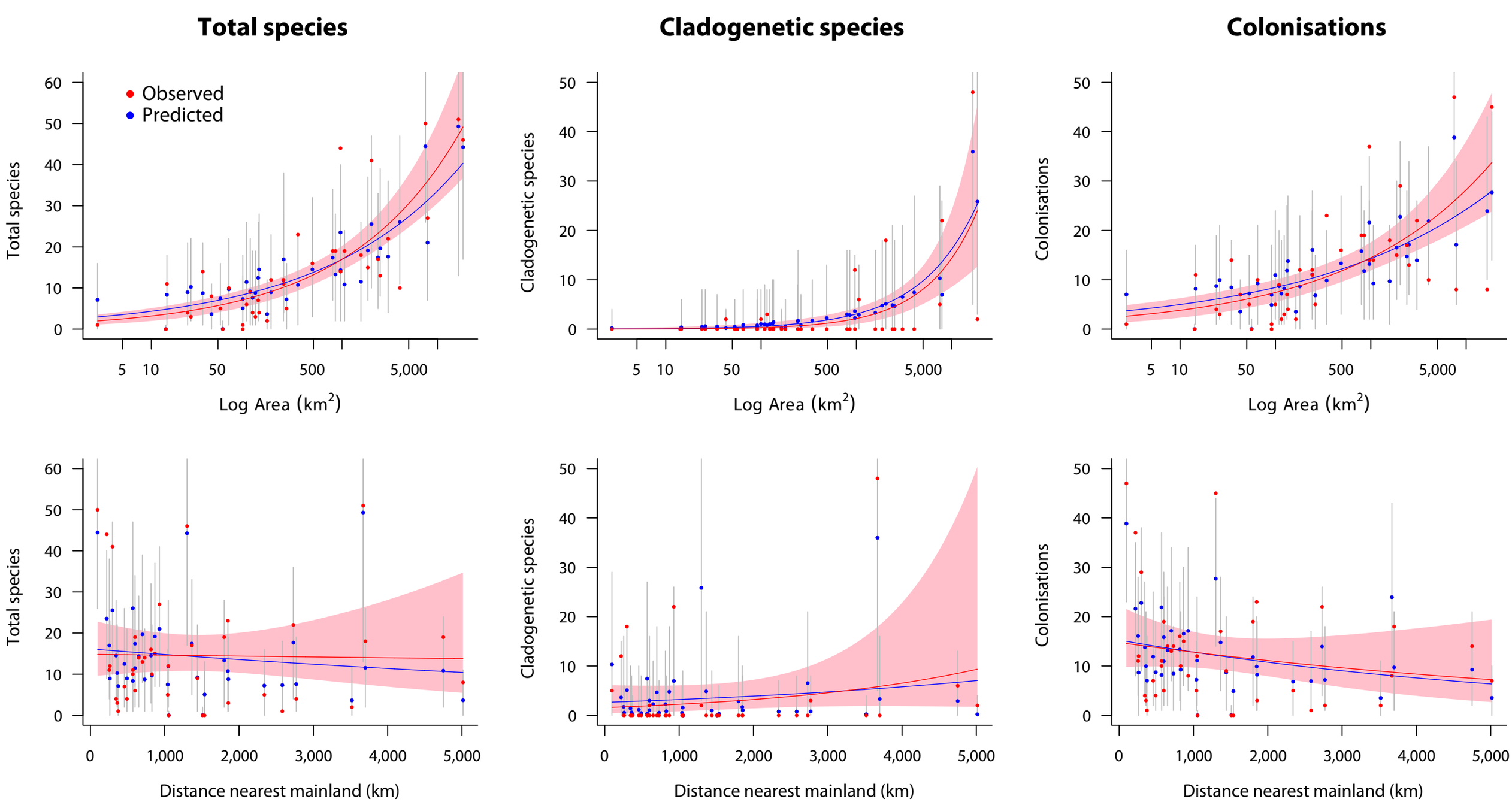
Extended Data Table 4 | Previously published dated trees used.

Extended Data Table 5 | Bootstrap of M14 and M19 models. ML estimates and 95% confidence intervals of the parameters of the two best models. Confidence intervals obtained from the bootstrap analyses. Parameter symbols explained in Supplementary Table 1.


1 Cocos (Keeling) 0|0
2 Chagos 0|0
3 Rapa Nui 2|0
4 Aldabra Group 12|91.7
5 Ascension 0|0
6 Pitcairn 8|28.6
7 Bermuda 5|40
8 Niue 5|60
9 Cocos 4|50
10 Chatham 14|92.9
11 Norfolk 14|78.6
12 Fernando de Noronha 3|33.3
13 Socorro 7|100
14 Galápagos 27|100
15 Gough 1|100
16 Society 18|38.9
17 Ogasawara 10|50
18 Reunion 13|69.2
19 Marquesas 19|50
20 Juan Fernández 6|100
21 Azores 17|76.5
22 Lord Howe 11|81.8
23 Guadalupe 11|90.9
24 Mauritius Isl. 15|60
25 Christmas Island 4|50
26 Samoa 22|72.7
27 Saint Helena 3|0
28 Comoros 41|96.5
29 Marianas 19|57.9
30 Rodrigues 9|44.4
31 Cape Verde 10|100
32 Tristan da Cunha 4|100
33 Madeira 19|89.5
34 Palau 16|62.5
35 Canary Islands 50|87.2
36 Selvagens 1|100
37 Hawaii 51|75
38 São Tomé e Príncipe 44|100
39 New Caledonia 46|82.2
40 Tonga 23|43.5
41 Seychelles (Inner) 12|91.7
Total species
0
10
20
30
40
50
With in-situ radiations
No extant in-situ radiations
60
50
40
30
20
10
0
Age (Myr)


Extinction rate (Myr^{-1})
0.5
1.0
1.5
2.0
2.5
Cladogenesis rate (Myr^{-1})
0.05
0.15
0.25
0.30
Far
Near
1
10
100
1,000
10,000
Area (km^2)
Colonisation rate (Myr^{-1})
0.000
0.010
0.020
0.030
Anagenesis rate (Myr^{-1})
0.0
0.5
1.0
1.5
0
1,000
2,000
3,000
4,000
5,000
Distance to mainland (km)


Colonisations
Total species
Cladogenetic species
Well estimated
Underestimated
Overestimated
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41


Total species
Cladogenetic species
Colonisations
Observed
Predicted
Total species
Cladogenetic species
Colonisations
Log Area (km²)
Distance nearest mainland (km)
0
10
20
30
40
50
60
5
10
50
500
5,000
1,000
2,000
3,000
4,000