



HAL
open science

A simple dynamic model explains the diversity of island birds worldwide

Luis Valente, Albert Phillimore, Martim Melo, Ben Warren, Sonya Clegg, Katja Havenstein, Ralph Tiedemann, Juan Carlos Illera, Christophe Thébaud, Tina Aschenbach, et al.

► To cite this version:

Luis Valente, Albert Phillimore, Martim Melo, Ben Warren, Sonya Clegg, et al.. A simple dynamic model explains the diversity of island birds worldwide. *Nature*, 2020, 579 (7797), pp.92-96. 10.1038/s41586-020-2022-5 . hal-02924282

HAL Id: hal-02924282

<https://hal.science/hal-02924282v1>

Submitted on 27 Aug 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A simple dynamic model explains island bird diversity worldwide

Luis Valente^{1,2,*}, Albert B. Phillimore³, Martim Melo^{4,5,6}, Ben H. Warren⁷, Sonya M. Clegg⁸,
Katja Havenstein², Ralph Tiedemann², Juan Carlos Illera⁹, Christophe Thébaud¹⁰, Tina
Aschenbach¹, Rampal S. Etienne¹¹

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

² Unit of Evolutionary Biology/Systematic Zoology, Institute of Biochemistry and Biology, University of
Potsdam, Karl-Liebknecht-Strasse 24–25, Haus 26, 14476 Potsdam, Germany

³ Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, EH9 3JT, UK

⁴ Museu de História Natural e da Ciência da Universidade do Porto, Praça Gomes Teixeira, 4099-002
Porto, Portugal

⁵ CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, InBio, Laboratório Associado,
Universidade do Porto, Campus Agrário de Vairão, 4485-661 Vairão, Portugal

⁶ FitzPatrick Institute, DST-NRF Centre of Excellence, University of Cape Town, Rondebosch 7701, Cape
Town, South Africa

⁷ Institut de Systématique, Evolution, Biodiversité (ISYEB), Muséum National d'Histoire Naturelle, CNRS,
Sorbonne Université, EPHE, UA, CP 51, 57 Rue Cuvier, 75005 Paris, France

⁸ Edward Grey Institute, Department of Zoology, University of Oxford, Oxford, OX1 3PS, United Kingdom

⁹ Research Unit of Biodiversity (UO-CSIC-PA), Oviedo University, 33600 Mieres, Asturias, Spain

¹⁰ Unité Mixte de Recherche 5174, CNRS-IRD-Paul Sabatier University, Toulouse, France

¹¹ Groningen Institute for Evolutionary Life Sciences, University of Groningen, P.O. Box 11103, Groningen
9700 CC, the Netherlands

*Corresponding author

26 Colonisation, speciation and extinction are dynamic processes that influence
27 global patterns of species richness¹⁻⁵. The theory of island biogeography predicts
28 that the contribution of these processes to the build-up of species diversity
29 depends on habitat area and isolation^{6,7}. Remarkably, to date there has been no
30 globally robust test of this fundamental prediction⁸, because neither the
31 appropriate data nor the analytical tools have been available. Here, we address
32 both deficiencies to reveal, for island birds, the empirical shape of the general
33 relationships that determine how colonisation, extinction and speciation rates
34 vary with island area and isolation. We present the first global molecular
35 phylogenetic dataset of times of colonisation and speciation of birds on islands,
36 based on the terrestrial avifaunas of 41 oceanic archipelagos worldwide,
37 including information on the 596 taxa on these archipelagos and their closest
38 mainland relatives. We apply novel methodology based on a dynamic model to
39 these data to estimate the sensitivity of island-specific rates of colonisation,
40 speciation and extinction to island features (area, isolation). Our simple eight-
41 parameter model predicts, with high explanatory power, several global
42 relationships that are fundamental to the origin and maintenance of species
43 assemblages: a decline of colonisation with isolation, a decline of extinction with
44 area, and an increase of speciation with area and isolation. Combining the
45 theoretical foundations of island biogeography^{6,7} with the temporal information
46 contained in molecular phylogenies⁹ proves a powerful approach to finally unveil
47 the fundamental relationships that govern variation in biodiversity across the
48 planet.

49 A key feature of the geographical unevenness of global diversity is the tendency for
50 some areas to harbour many more species than others^{6,7}. Uncovering the drivers and
51 regulators of spatial differences in diversity of simple systems such as islands is a
52 crucial step towards understanding the global distribution of species richness. The two
53 most prominent biodiversity patterns in fragmented or isolated environments
54 worldwide are the increase of species richness with area and the decline of richness
55 with isolation^{7,10-13}. In their theory of island biogeography, MacArthur and Wilson
56 proposed how the processes of colonisation and extinction could explain these
57 patterns^{6,7}. They argued that the rates of these processes are determined by the
58 geographic context: colonisation decreases with isolation and extinction decreases with
59 area^{6,7}. They also suggested that rates of formation of island endemic species via *in situ*
60 speciation increase with island isolation and area⁷. Despite an abundance of studies
61 over five decades supporting the general patterns predicted by MacArthur and
62 Wilson^{2,14-17}, tests of predictions regarding the dependence of the underlying processes
63 – colonisation, speciation and extinction – on island geographic context (area and
64 isolation) are few in number, and are either restricted in temporal, geographic, or
65 taxonomic scope^{5,18,19} or seek to infer speciation rates in the absence of data on the
66 relationships among species^{2,15}. As a result, there has been no robust and powerful test
67 of MacArthur and Wilson’s predictions on a global scale, and the effect of area and
68 isolation on biogeographical processes acting on macro-evolutionary time scales
69 remains largely unexplored.

70 Here we expand on approaches that leverage the information in time-calibrated
71 molecular phylogenies of insular species^{1,9,20,21} to determine how the processes of
72 colonisation, speciation and extinction are influenced by area and isolation. The
73 dynamic stochastic model DAISIE⁹ (Dynamic Assembly of Islands through Speciation,
74 Immigration, and Extinction) can accurately estimate maximum likelihood rates of
75 colonisation, extinction and speciation rates (CES rates) from branching times
76 (colonisation times and any *in situ* diversification events) and endemism status of
77 species resulting from one or multiple independent colonisations of a given island
78 system (e.g. all native terrestrial birds on an archipelago)⁹. This method can also detect
79 the presence or absence of diversity-dependence in rates of colonisation and speciation,
80 by estimating a carrying capacity (upper bound to the number of species in an island
81 system). Here we extend DAISIE to estimate, for the first time, the hyperparameters

82 (parameters that determine downstream parameters) that control the shape of the
83 relationships between CES rates and the area and isolation of islands worldwide.

84 Accurate estimation of fundamental island biogeographic relationships requires
85 suitable data from many archipelagos, but divergence-dated phylogenies of complete
86 communities on islands remain scarce. Hence, we produced new dated molecular
87 phylogenies for the terrestrial avifaunas of 41 archipelagos worldwide. By ‘archipelago’
88 we refer to both true archipelagos (composed of multiple islands) and isolated insular
89 units consisting of single islands (e.g. Saint Helena). For each archipelago we compiled
90 avian taxon lists (excluding introduced, marine, migratory, and aquatic species, as well
91 as birds of prey, rails and nocturnal birds, see Methods) and collected physical data (Fig.
92 1, Supplementary Data 1 and 2). We use archipelagos as our insular unit, because the
93 high dispersal abilities of birds within archipelagos imply that for birds, archipelagos
94 can be considered equivalent to single islands for less dispersive taxa²², and because
95 archipelagos constitute the most appropriate spatiotemporal unit for framing analyses
96 of biodiversity patterns at a large scale^{2,23}. We extracted colonisation and speciation
97 times for each archipelago from the phylogenetic trees, producing a ‘global dataset’ for
98 the 41 archipelagos, which includes each archipelago’s complete extant avifauna, plus
99 all species known to have become extinct due to anthropogenic causes. The dataset
100 comprises 596 insular taxa from 491 species. The phylogenies revealed a total of 502
101 archipelago colonisation events and 26 independent in-situ ‘radiations’ (cases where
102 diversification has occurred within an archipelago) ranging in size from 2 to 33 species
103 (the Hawaiian honeycreepers being the largest clade). The distribution of colonisation
104 times is summarised in Fig. 1 and the full dataset is given in Supplementary Data 1.

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

115 cladogenesis and carrying capacity^{3,5,7}, and a negative effect on extinction rates^{7,24}.
116 Archipelago isolation is hypothesised to reduce colonisation rates⁶ and elevate
117 anagenesis rates²⁵. Models including or excluding diversity-dependence in rates of
118 colonisation and cladogenesis⁹ (i.e. estimating a carrying capacity parameter) were
119 compared. We also considered a set of *post hoc* models with alternative shapes for the
120 relationships (*post hoc* power and *post hoc* sigmoid models, see Methods,
121 Supplementary Information Table 1).

122 We fitted a set of 28 candidate models to the global dataset using maximum
123 likelihood (Supplementary Information Table 2). The shape of the relationship of CES
124 rates with area and distance for the two best models is shown in Fig. 2. Under the
125 preferred *a priori* model (lowest value of Bayesian Information Criterion (BIC); M14,
126 eight parameters) colonisation rates decline with archipelago isolation (exponent of the
127 power law = -0.25 (95% confidence interval = -0.17 - -0.34)) and extinction rate
128 decreases with area (scaling = -0.15 (-0.11 - -0.18)). Rates of cladogenesis increase
129 with area (scaling = 0.26 (0.13 - 0.37)), whilst anagenesis increases with isolation
130 (scaling = 0.42 (0.24 - 0.61)). The preferred *post hoc* model (M19, eight parameters) was
131 also the preferred model overall and differs qualitatively from the preferred *a priori*
132 model M14 only in the cladogenesis function. In M14 cladogenesis is solely a function of
133 area, whereas in M19 cladogenesis depends interactively and positively on both area
134 and distance from the nearest mainland, such that the cladogenesis-area relationship is
135 steeper for more isolated archipelagos (Fig. 2 and Extended Data Fig. 1). In addition, we
136 found no evidence for diversity-dependence, as K was estimated to be much larger than
137 the number of species on the island and models without a carrying capacity parameter
138 (no upper bound to diversity), such as M14 and M19, performed better than models
139 including one (Supplementary Information Table 2). We also tested whether the
140 inclusion of a combination of true archipelagos and single islands in our dataset could
141 have affected our results, for example if opportunities for allopatric speciation are
142 higher when an area is subdivided into multiple islands²⁶. We repeated analyses
143 excluding single island units and found that the same model (M19) is preferred with
144 similar parameter estimates. Hence, we discuss only the results for the main dataset
145 (including both single islands and true archipelagos). Our results are robust to
146 uncertainty in colonisation and branching times (see 'Sensitivity to alternative
147 divergence times and tree topologies section').

148 A parametric bootstrap analysis of the two preferred models (M14 and M19)
149 demonstrated that the method is able to recover hyperparameters with precision and
150 little bias (Extended Data Figs. 2 and 3). In order to test the significance of the
151 relationships between area, isolation and CES rates, we conducted a randomization test
152 on the global dataset with reshuffled areas and distances. This test estimated the
153 exponent hyperparameters as zero in most reshuffled cases (i.e. no effect of area or
154 isolation detected; Extended Data Fig. 4), confirming that the data are informative.

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

168 Our approach reveals the empirical shape of fundamental biogeographic
169 relationships that have hitherto largely evaded estimation. In agreement with recent
170 studies^{2,27}, we found strong evidence for a decline of rates of colonisation with isolation
171 and of rates of extinction with area, confirming two of the key assumptions of island
172 biogeography theory⁶. The colonisation-isolation effect was detected despite the decline
173 of avian richness with distance from the nearest mainland in our empirical data (Fig. 4)
174 not being as pronounced as in other less mobile taxa^{4,10}, revealing isolation to be a clear
175 determinant of probability of immigration and successful establishment of populations
176 even in a highly dispersive group such as birds. The extinction-area relationship has
177 been a fundamental empirical generalization in conservation theory (for example for
178 the design of protected areas²⁸) and yet this is the first time the shape of this
179 dependence is characterized at the global spatial scale and macro-evolutionary time
180 scale.

181 We provide novel insights into the scaling of speciation with area and isolation.
182 Contrary to previous work on within-island speciation, which suggested the existence of
183 an area below which cladogenesis does not take place on single islands⁵, we do not find
184 evidence for such an area threshold at the archipelago level. In addition, our *post hoc*
185 finding that rates of cladogenesis increase through an interactive effect of both island
186 size and distance from the nearest mainland (Fig. 2 and Extended Data Fig. 1) provides a
187 mechanism that limits radiations to archipelagos that are both large and remote^{25,29}.
188 Why this interaction exists requires further investigation, but one possibility is that
189 unsaturated niche space provides greater opportunities for diversification²⁹. While we
190 do not find any evidence that diversity of a particular lineage feeds back negatively on
191 rates of cladogenesis and colonisation of descendants of the ancestral lineage (as in
192 Valente et al.³⁰), it remains possible that diversity dependence acts at the archipelago
193 scale on multiple lineages³¹. In addition to the effects of physical features on
194 cladogenesis, we found that rates of anagenesis increase with island isolation. While
195 impressive insular radiations tend to receive the most attention from evolutionary
196 biologists (e.g. Darwin's finches or Hawaiian honeycreepers), our phylogenies revealed
197 that the majority of endemic birds in our dataset in fact display an anagenetic pattern
198 (at the time of human arrival 231 of 350 endemic species had no extant sister taxa on
199 the archipelago and there were only 26 extant in situ radiations). The positive effect of
200 archipelago isolation on rates of anagenesis that we estimate suggests this fundamental
201 but overlooked process is impeded by high levels of movement between island and
202 mainland populations.

203 A variety of global patterns of biodiversity have been described – from small
204 islands and lakes, up to biomes and continents^{4,32} - but the processes underpinning
205 them remain little explored. Our simulations using parameters estimated from data
206 were able to reproduce well-known global patterns of island biogeography across 41
207 archipelagos (Fig. 4). This advances our understanding of macro-scale biology, by
208 providing missing links between local process, environment and global patterns. Over
209 half a century since the seminal work of MacArthur & Wilson⁶, we now have the data
210 and tools to go beyond diversity patterns, enabling us to quantify community-level
211 processes that have long been elusive. A next step will be to examine the extent to
212 which the parameters and shapes of relationships we estimate generalise to habitat-
213 islands and continental regions, and to other less vagile taxa. Crucially, our inference

214 approach can contribute to reduce uncertainty in estimating key parameters for
215 biodiversity conservation worldwide, such as extinction rates in a context of habitat loss
216 and fragmentation and time to recovery to equilibrium from perturbations.

217

218 **Acknowledgments:**

219 **(As these are quite long we can move some to the Methods section if requested)**

220 We thank the skilled guides and field assistants who helped with sample collection in
221 the field. For comments on the manuscript: Alex Pigot, Tom Matthews and two
222 anonymous reviewers. For providing samples or DNA sequences: Nancy Bunbury
223 (Seychelles Islands Foundation) who organized sample loans of Aldabra island; Janske
224 van de Crommenacke, Jim Groombridge, Hazel Jackson. For sharing data on extinct
225 species: Josep Antoni Alcover, Juan Carlos Rando, Ferran Sayol and Søren Faurby. For
226 support/advice at the MfN: Kristina and Thomas von Rintelen. For permission to use
227 photographs or illustrations: Cláudia Baeta, Martijn Hammers, Dubi Shapiro, Juan
228 Varela and Pedro Cascão. For expertise on island geological ages: Paul Hearty, Robert
229 Stern and Mark Reagan. For providing phylogenetic data: Alice Cibois, Jimmy McGuire,
230 Heather Lerner, Petter Marki, Borja Milá, Guillermo Friis, Jérôme Fuchs, John P.
231 Dumbacher and Ore Carmi. For permission to obtain or use existing samples, and for
232 logistic support: São Tomé e Príncipe: Arlindo Carvalho and the Department of the
233 Environment; Equatorial Guinea: José Obiang, Noélia Calvo, the Universidad Nacional de
234 Guinea Ecuatorial for Bioko and Annobón samples; Seychelles: the Ministry of
235 Environment, Energy and Climate Change of the Republic of Seychelles, the Seychelles
236 Bureau of Standards, BirdLife Seychelles; Comoros: Centre National de Documentation
237 et de Recherche Scientifique (Grande Comore & Anjouan), Action Comores, Direction de
238 l'Agriculture et de la Forêt (Mayotte); Madagascar: Ministère des Eaux et Forêts
239 (Madagascar), the Madagascar Institute pour la Conservation des Ecosystemes
240 Tropicaux; Mauritius: Mauritius National Parks and Conservation Service, Mauritius
241 Wildlife Foundation; New Caledonia: Olivier Hébert, Waifite Waheoneme, Nicholas
242 Clark, the Direction de L'Environnement (South Province), Direction du Développement
243 Economique (Loyalty Islands Province), local chiefs and landowners; Morocco:
244 Moroccan Environment Ministry; Cape Verde: Cape Verde Agriculture and Environment
245 Ministry; Cameroon: Francis Njie and the Limbe Botanical and Zoological Garden;
246 Gabon: Station de Recherche de l'IRET at Ipassa-Makokou; Angola: Fernanda Lages

247 (ISCED-Huía); Spain: the regional governments of Andalucía and the Canary Islands;
248 Portugal: regional governments of Madeira and the Azores. Museum samples:
249 Department of Ornithology and Mammalogy of the California Academy of Sciences
250 (Laura Wilkinson & Maureen Flannery) for loaning Galápagos samples; Natural History
251 Museum at Tring (Mark Adams) for loaning Comoros samples; the Stuttgart State
252 Museum of Natural History for loaning stonechat samples from Madagascar. Sebastian
253 Block assisted with cluster analyses at the Museum für Naturkunde. The Center for
254 Information Technology of the University of Groningen provided support and access to
255 the Peregrine high-performance computing cluster. L.V. was funded by the German
256 Science Foundation (DFG Research grant VA 1102/1-1), the Alexander von Humboldt
257 Foundation, the Brandenburg Postdoc Prize and by a VIDI grant from the Netherlands
258 Organisation for Scientific Research (NWO); R.S.E. by a NOW VICI grant; M.M. by the
259 Portuguese Science and Technology Foundation (Post-doctoral grant:
260 SFRH/BPD/100614/2014); S.M.C by the National Geographic Society (CRE grant #
261 9383-13); J.C.I. by a GRUPIN research grant from the Regional Government of Asturias
262 (Ref.: IDI/2018/000151);); C.T. by the 'Laboratoire d'Excellence' TULIP (ANR-10-
263 LABX-41).

264

265 **Author contribution**

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

273

274 **Data availability statement**

275

276 Sequence data that support the findings of this study have been deposited in GenBank
277 with the accession codes: MH307408- MH307656. The phylogenetic trees and DAISIE
278 objects generated and/or analysed during the current study have been deposited in

279 Mendeleev (trees: XXXX, DAISIE objects: XXXX Under embargo until publication. The 11
280 previously published trees are available upon request.

281

282 Code availability

283

284 All custom computer code used for this study is freely available in the DAISIE R package
285 (Github).

286

287 References

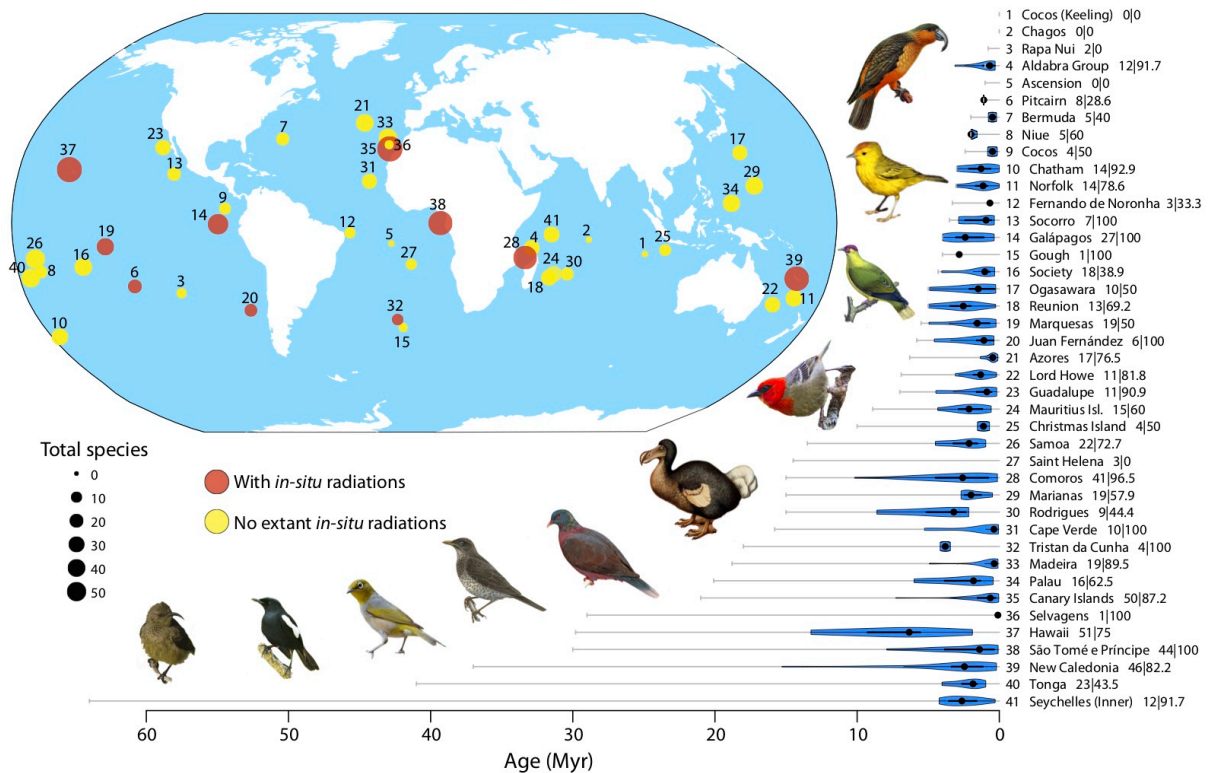
- 288 1. Ricklefs, R. E. & Bermingham, E. Nonequilibrium diversity dynamics of the Lesser Antillean
289 avifauna. *Science (80-.)*. **294**, 1522–1524 (2001).
- 290 2. Triantis, K. A., Economo, E. P., Guilhaumon, F. & Ricklefs, R. E. Diversity regulation at macro-scales:
291 species richness on oceanic archipelagos. *Glob. Ecol. Biogeogr.* **24**, 594–605 (2015).
- 292 3. Whittaker, R. J., Triantis, K. A. & Ladle, R. J. A general dynamic theory of oceanic island
293 biogeography. *J. Biogeogr.* **35**, 977–994 (2008).
- 294 4. Kreft, H., Jetz, W., Mutke, J., Kier, G. & Barthlott, W. Global diversity of island floras from a
295 macroecological perspective. *Ecol. Lett.* **11**, 116–127 (2008).
- 296 5. Losos, J. B. & Schluter, D. Analysis of an evolutionary species-area relationship. *Nature* **408**, 847–
297 850 (2000).
- 298 6. MacArthur, R. H. & Wilson, E. O. An equilibrium theory of insular zoogeography. *Evolution (N. Y.)*.
299 **17**, 373–387 (1963).
- 300 7. MacArthur, R. H. & Wilson, E. O. *The Theory of Island Biogeography*. (Princeton University Press,
301 1967).
- 302 8. Warren, B. H. *et al.* Islands as model systems in ecology and evolution: Prospects fifty years after
303 MacArthur-Wilson. *Ecol. Lett.* **18**, 200–217 (2015).
- 304 9. Valente, L. M., Phillimore, A. B. & Etienne, R. S. Equilibrium and non-equilibrium dynamics
305 simultaneously operate in the Galápagos islands. *Ecol. Lett.* **18**, 844–852 (2015).
- 306 10. Lomolino, M. V. Species-area and species-distance relationships of terrestrial mammals in the
307 Thousand Island Region. *Oecologia* **54**, 72–75 (1982).
- 308 11. Diamond, J. M. Biogeographic kinetics: estimation of relaxation times for avifaunas of Southwest
309 Pacific islands. *Proc. Natl. Acad. Sci.* **69**, 3199–3203 (1972).
- 310 12. Whittaker, R. J. & Fernandez-Palacios, J. M. *Island Biogeography: Ecology, Evolution, and*
311 *Conservation*. (Oxford University Press, 2007).
- 312 13. Matthews, T. J., Rigal, F., Triantis, K. A. & Whittaker, R. J. A global model of island species–area
313 relationships. *Proc. Natl. Acad. Sci.* 201818190 (2019).
- 314 14. Weigelt, P., Steinbauer, M. J., Cabral, J. S. & Kreft, H. Late Quaternary climate change shapes island
315 biodiversity. *Nature* **532**, 99–102 (2016).
- 316 15. Lim, J. Y. & Marshall, C. R. The true tempo of evolutionary radiation and decline revealed on the
317 Hawaiian archipelago. *Nature* **543**, 710–713 (2017).
- 318 16. Cabral, J. S., Weigelt, P., Kissling, W. D. & Kreft, H. Biogeographic, climatic and spatial drivers
319 differentially affect α -, β - and γ -diversities on oceanic archipelagos. *Proc. R. Soc. London B Biol. Sci.*
320 **281**, 20133246 (2014).
- 321 17. Matthews, T. J., Guilhaumon, F., Triantis, K. A., Borregaard, M. K. & Whittaker, R. J. On the form of
322 species–area relationships in habitat islands and true islands. *Glob. Ecol. Biogeogr.* **25**, 847–858
323 (2016).
- 324 18. Simberloff, D. S. & Wilson, E. O. Experimental zoogeography of islands: the colonization of empty
325 islands. *Ecology* **50**, 278–296 (1969).
- 326 19. Russell, G. J., Diamond, J. M., Reed, T. M. & Pimm, S. L. Breeding birds on small islands: island
327 biogeography or optimal foraging? *J. Anim. Ecol.* **75**, 324–339 (2006).
- 328 20. Rabosky, D. L. & Glor, R. E. Equilibrium speciation dynamics in a model adaptive radiation of
329 island lizards. *Proc. Natl. Acad. Sci.* **107**, 22178–22183 (2010).
- 330 21. Emerson, B. C. & Gillespie, R. G. Phylogenetic analysis of community assembly and structure over

- 331 space and time. *Trends Ecol. Evol.* **23**, 619–30 (2008).
- 332 22. Kisel, Y. & Barraclough, T. G. Speciation has a spatial scale that depends on levels of gene flow. *Am*
333 *Nat* **175**, 316–334 (2010).
- 334 23. Triantis, K., Whittaker, R. J., Fernández-Palacios, J. M. & Geist, D. J. Oceanic archipelagos: a
335 perspective on the geodynamics and biogeography of the World’s smallest biotic provinces. *Front.*
336 *Biogeogr.* **8**, e29605 (2016).
- 337 24. Ricklefs, R. E. & Lovette, I. J. The roles of island area per se and habitat diversity in the species-area
338 relationships of four Lesser Antillean faunal groups. *J. Anim. Ecol.* **68**, 1142–1160 (1999).
- 339 25. Rosindell, J. & Phillimore, A. B. A unified model of island biogeography sheds light on the zone of
340 radiation. *Ecol. Lett.* **14**, 552–560 (2011).
- 341 26. Losos, J. B. & Ricklefs, R. E. Adaptation and diversification on islands. *Nature* **457**, 830–836 (2009).
- 342 27. Keil, P. *et al.* Extinction rate has a complex and non-linear relationship with area. *bioRxiv* 81489
343 (2016). doi:10.1101/081489
- 344 28. Wilcox, B. A. & Murphy, D. D. Conservation strategy: the effects of fragmentation on extinction. *Am.*
345 *Nat.* **125**, 879–887 (1985).
- 346 29. Gillespie, R. G. & Baldwin, B. G. Island biogeography of remote archipelagos. in *The Theory of Island*
347 *Biogeography Revisited* (eds. Losos, J. & Ricklefs, R. E.) 358–387 (Princeton University Press,
348 2010).
- 349 30. Valente, L. *et al.* Equilibrium bird species diversity in Atlantic islands. *Curr. Biol.* **27**, 1660–1666
350 (2017).
- 351 31. Borregaard, M. K., Matthews, T. J. & Whittaker, R. J. The general dynamic model: towards a unified
352 theory of island biogeography? *Glob. Ecol. Biogeogr.* **25**, 805–816 (2015).
- 353 32. Gaston, K. J. Global patterns in biodiversity. *Nature* **405**, 220–227 (2000).
- 354 33. Plummer, P. S. & Belle, E. R. Mesozoic tectono-stratigraphic evolution of the Seychelles
355 microcontinent. *Sediment. Geol.* **96**, 73–91 (1995).
- 356 34. Norder, S. J. *et al.* Beyond the Last Glacial Maximum: Island endemism is best explained by long-
357 lasting archipelago configurations. *Glob. Ecol. Biogeogr.* (2018).
358 doi:https://doi.org/10.1111/geb.12835
- 359 35. Thomson, J. & Walton, A. Redetermination of chronology of alibaba atoll by 230Th/ 234U dating.
360 *Nature* **240**, 145–146 (1972).
- 361 36. Price, J. P. & Clague, D. A. How old is the Hawaiian biota? Geology and phylogeny suggest recent
362 divergence. *Proc. Biol. Sci.* **269**, 2429–35 (2002).
- 363 37. *Handbook of the Birds of the World Alive*. (Lynx Edicions, 2018).
- 364 38. Valente, L., Etienne, R. S. & Dávalos, L. M. Recent extinctions disturb path to equilibrium diversity
365 in Caribbean bats. *Nat. Ecol. Evol.* **1**, 26 (2017).
- 366 39. Steadman, D. W. *Extinction and biogeography of tropical Pacific birds*. (University of Chicago Press,
367 2006).
- 368 40. Cheke, A. & Hume, J. P. *Lost land of the Dodo: The ecological history of Mauritius, Réunion and*
369 *Rodrigues*. (Bloomsbury Publishing, 2010).
- 370 41. Lerner, H. R. L., Meyer, M., James, H. F., Hofreiter, M. & Fleischer, R. C. Multilocus resolution of
371 phylogeny and timescale in the extant adaptive radiation of Hawaiian honeycreepers. *Curr. Biol.*
372 **21**, 1838–44 (2011).
- 373 42. Rando, J. C., Pieper, H., Olson, S. L., Pereira, F. & Alcover, J. A. A new extinct species of large
374 bullfinch (Aves: Fringillidae: Pyrrhula) from Graciosa Island (Azores, North Atlantic Ocean).
375 *Zootaxa* **4282**, 567–583 (2017).
- 376 43. Illera, J. C., Rando, J. C., Richardson, D. S. & Emerson, B. C. Age, origins and extinctions of the
377 avifauna of Macaronesia: a synthesis of phylogenetic and fossil information. *Quat. Sci. Rev.* **50**, 14–
378 22 (2012).
- 379 44. Hume, J. P., Martill, D. & Hing, R. A terrestrial vertebrate palaeontological review of Aldabra Atoll,
380 Aldabra Group, Seychelles. *PLoS One* **13**, e0192675 (2018).
- 381 45. Cheke, A. S. Extinct birds of the Mascarenes and Seychelles—a review of the causes of extinction in
382 the light of an important new publication on extinct birds. *Phelsuma* **21**, 4–19 (2013).
- 383 46. Hume, J. P. & Walters, M. *Extinct birds*. (A&C Black, 2012).
- 384 47. Kearse, M. *et al.* Geneious Basic: an integrated and extendable desktop software platform for the
385 organization and analysis of sequence data. *Bioinformatics* **28**, 1647–9 (2012).
- 386 48. Bouckaert, R. *et al.* BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput.*
387 *Biol.* **10**, e1003537 (2014).
- 388 49. Posada, D. jModelTest: Phylogenetic Model Averaging. *Mol Biol Evol* **25**, 1253–1256 (2008).
- 389 50. Weir, J. T. & Schluter, D. Calibrating the avian molecular clock. *Mol. Ecol.* **17**, 2321–2328 (2008).

- 390 51. Field, D. J. *et al.* Timing the extant avian radiation: the rise of modern birds, and the importance of
391 modeling molecular rate variation. *PeerJ Prepr.* **7**, e27521v1 (2019).
- 392 52. Cicero, C. & Johnson, N. K. Higher-level phylogeny of New World vireos (Aves: Vireonidae) based
393 on sequences of multiple mitochondrial DNA genes. *Mol. Phylogenet. Evol.* **20**, 27–40 (2001).
- 394 53. Valente, L., Phillimore, A. & Etienne, R. S. Using molecular phylogenies in island biogeography: it's
395 about time. *Ecography (Cop.)*. **41**, 1684–1686 (2018).
- 396 54. Rabosky, D. L. Extinction rates should not be estimated from molecular phylogenies. *Evolution (N.*
397 *Y)*. **64**, 1816–1824 (2010).
- 398 55. Nee, S., May, R. M. & Harvey, P. H. The reconstructed evolutionary process. *Philos. Trans. R. Soc.*
399 *Lond. B. Biol. Sci.* **344**, 305–11 (1994).
- 400 56. Etienne, R. S. *et al.* Diversity-dependence brings molecular phylogenies closer to agreement with
401 the fossil record. *Proc. R. Soc. B Biol. Sci.* **279**, 1300–1309 (2012).
- 402 57. Ogilvie, H. A., Heled, J., Xie, D. & Drummond, A. J. Computational Performance and Statistical
403 Accuracy of *BEAST and Comparisons with Other Methods. *Syst. Biol.* **65**, 381–396 (2016).
- 404 58. Maddison, W. P. & Knowles, L. L. Inferring Phylogeny Despite Incomplete Lineage Sorting. *Syst.*
405 *Biol.* **55**, 21–30 (2006).
- 406 59. Lemmon, A. R., Brown, J. M., Stanger-Hall, K. & Lemmon, E. M. The Effect of Ambiguous Data on
407 Phylogenetic Estimates Obtained by Maximum Likelihood and Bayesian Inference. *Syst. Biol.* **58**,
408 130–145 (2009).
- 409 60. Weigelt, P. & Kreft, H. Quantifying island isolation—insights from global patterns of insular plant
410 species richness. *Ecography (Cop.)*. **36**, 417–429 (2013).
- 411 61. Nielson, D. L. & Sibbett, B. S. Geology of ascension Island, South Atlantic Ocean. *Geothermics* **25**,
412 427–448 (1996).
- 413 62. Ramalho, R. S. *et al.* Emergence and evolution of Santa Maria Island (Azores)—The conundrum of
414 uplifted islands revisited. *Geol. Soc. Am. Bull.* **129**, 372–390 (2017).
- 415 63. Hearty, P. J. & Olson, S. L. Geochronology, biostratigraphy, and changing shell morphology in the
416 land snail subgenus *Poecilozonites* during the Quaternary of Bermuda. *Palaeogeogr.*
417 *Palaeoclimatol. Palaeoecol.* **293**, 9–29 (2010).
- 418 64. Carracedo, J. C. & Troll, V. R. *The Geology of the Canary Islands*. (Elsevier, 2016).
- 419 65. Ramalho, R. *Building the Cape Verde Islands*. (Springer, 2011).
- 420 66. Eisenhauer, A., Heiss, G. A., Sheppard, C. R. C. & Dullo, W. C. Reef and island formation and Late
421 Holocene sea-level changes in the Chagos islands. *Ecol. Chagos Archipel.* 21–31 (1999).
- 422 67. Campbell, H. J. Fauna and flora of the Chatham Islands: less than 4 my old. *Geol. genes* **97**, 15–16
423 (1998).
- 424 68. Bullough, F. History and Geology of Christmas Island. *Geological Society of London blog*
425 <https://blog.geolsoc.org.uk/2013/12/18/door-18-his> (2013). Available at:
426 <https://blog.geolsoc.org.uk/2013/12/18/door-18-history-and-geology-of-christmas-island/>.
- 427 69. Castillo, P. *et al.* Anomalously young volcanoes on old hot-spot traces: I. Geology and petrology of
428 Cocos Island. *Geol. Soc. Am. Bull.* **100**, 1400–1414 (1988).
- 429 70. Woodroffe, C. D., Veeh, H. H., Falkland, A. C., McLean, R. F. & Wallensky, E. Last interglacial reef and
430 subsidence of the Cocos (Keeling) Islands, Indian Ocean. *Mar. Geol.* **96**, 137–143 (1991).
- 431 71. Nougier, J., Cantagrel, J. M. & Karche, J. P. The Comores archipelago in the western Indian Ocean:
432 volcanology, geochronology and geodynamic setting. *J. African Earth Sci.* **5**, 135–144 (1986).
- 433 72. Almeida, F. The Fernando de Noronha archipelago. in *Sítios Geológicos e Paleontológicos do Brasil*
434 (ed. Schobbenhaus, C.; Campos, D. A.; Queiroz, E. T.; Winge, M.; Berbert-Born, M.)
435 <http://www.unb.br/ig/sigep/sitio066/sitio066englis> (2000).
- 436 73. Ali, J. R. & Aitchison, J. C. Exploring the combined role of eustasy and oceanic island thermal
437 subsidence in shaping biodiversity on the Galápagos. *J. Biogeogr.* **41**, 1227–1241 (2014).
- 438 74. Ryan, P. G. Tristan da Cunha and Gough Island. in *Encyclopedia of Islands* (eds. Gillespie, R. &
439 Clague, D.) 929–932 (University of California Press, 2009).
- 440 75. Batiza, R. Petrology and chemistry of Guadalupe Island: An alkalic seamount on a fossil ridge crest.
441 *Geology* **5**, 760–764 (1977).
- 442 76. Stuessy, T. F., Foland, K. A., Sutter, J. F., Sanders, R. W. & O., M. S. Botanical and geological
443 significance of potassium-argon dates from the Juan Fernandez islands. *Science (80-.)*. **225**, 49–51
444 (1984).
- 445 77. McDougall, I., Embleton, B. J. J. & Stone, D. B. Origin and evolution of Lord Howe Island, Southwest
446 Pacific Ocean. *J. Geol. Soc. Aust.* **28**, 155–176 (1981).
- 447 78. Mata, J. *et al.* O arquipélago da Madeira - Geografia de Portugal. in *Geologia de Portugal* (eds. Dias,
448 R., Araujo, A., Terrinha, P. & Kullberg, J.) **2**, 691–746 (Escolar Editora, 2013).

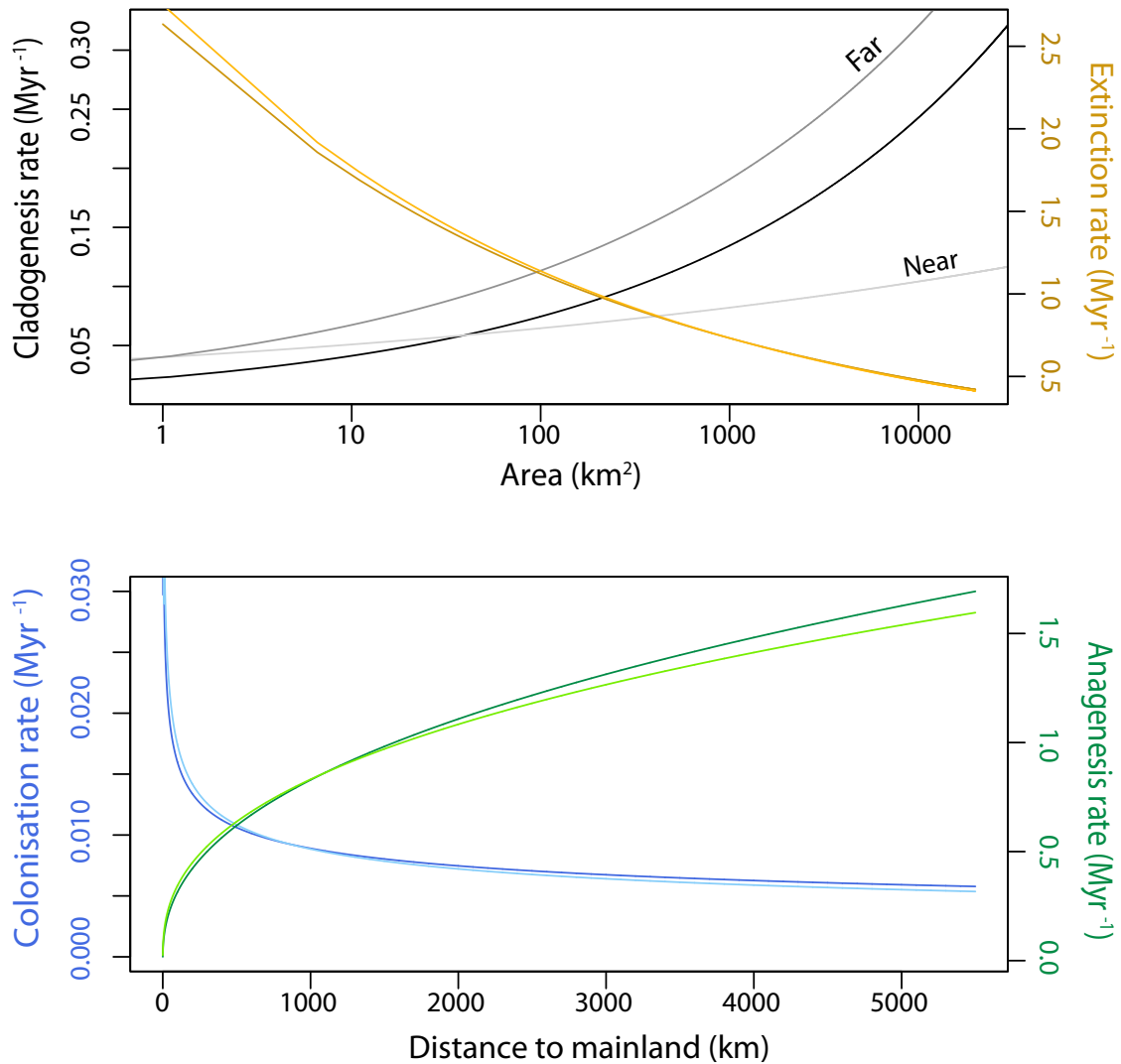
- 449 79. Guille, G. *et al.* Les marques (Polynésie Françaises): un archipel intraocéanique atypique. *Géologie la Fr.* **2**, 5–36 (2002).
- 450
- 451 80. Montaggioni, L. & Nativel, P. *La Reunion, Ile Maurice. Géologie et aperçus biologiques, plantes et*
- 452 *animaux.* (Masson, 1988).
- 453 81. Grandcolas, P. *et al.* New Caledonia: a very old Darwinian island? *Philos. Trans. R. Soc. B Biol. Sci.*
- 454 **363**, 3309–3317 (2008).
- 455 82. Anthoni, J. Geography and geology of Niue. <http://www.seafriends.org.nz/niue/geo.htm> (2005).
- 456 Available at: <http://www.seafriends.org.nz/niue/geo.htm>.
- 457 83. Jones, J. G. & McDougall, I. Geological history of Norfolk and Philip islands, southwest Pacific ocean.
- 458 *J. Geol. Soc. Aust.* **20**, 240–254 (1973).
- 459 84. Suzuki, M., Taisuke, S. & Hideo, T. *Nomination of the Ogasawara Islands for inscription on the World*
- 460 *Heritage List.* (Government of Japan, 2010).
- 461 85. Neall, V. E. & Trewick, S. A. The age and origin of the Pacific islands: a geological overview. *Philos.*
- 462 *Trans. R. Soc. B Biol. Sci.* **363**, 3293–3308 (2008).
- 463 86. Hekinian, R. *et al.* The Pitcairn hotspot in the South Pacific: Distribution and composition of
- 464 submarine volcanic sequences. *J. Volcanol. Geotherm. Res.* **121**, 219–245 (2003).
- 465 87. Vezzoli, L. & Acocella, V. Easter Island, SE Pacific: An end-member type of hotspot volcanism. *Bull.*
- 466 *Geol. Soc. Am.* **121**, 869–886 (2009).
- 467 88. Gillot, P.-Y., Lefèvre, J.-C. & Nativel, P.-E. Model for the structural evolution of the volcanoes of
- 468 Réunion Island. *Earth Planet. Sci. Lett.* **122**, 291–302 (1994).
- 469 89. Safford, R. & Hawkins, F. *The birds of Africa: Volume VIII: The Malagasy Region: Madagascar,*
- 470 *Seychelles, Comoros, Mascarenes.* **8**, (A&C Black, 2013).
- 471 90. Baker, I., Gale, N. H. & Simons, J. Geochronology of the St Helena volcanoes. *Nature* **215**, 1451–
- 472 1456 (1967).
- 473 91. 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
- 474 Publications, 1985).
- 475
- 476 92. Lee, D. C., Halliday, A. N., Fitton, J. G. & Poli, G. Isotopic variations with distance and time in the
- 477 volcanic islands of the Cameroon line: evidence for a mantle plume origin. *Earth Planet. Sci. Lett.*
- 478 **123**, 119–138 (1994).
- 479 93. Geldmacher, J., Hoernle, K., Van Den Bogaard, P., Zankl, G. & Garbe-Schönberg, D. Earlier history of
- 480 the ≥70-Ma-old Canary hotspot based on the temporal and geochemical evolution of the Selvagen
- 481 Archipelago and neighboring seamounts in the Eastern North Atlantic. *J. Volcanol. Geotherm. Res.*
- 482 **111**, 55–87 (2001).
- 483 94. Clouard, V. & Bonneville, A. Ages of seamounts, islands, and plateaus on the Pacific plate. in *Plates,*
- 484 *plumes and paradigms* (eds. Foulger, G. R., Natland, J. H., Presnall, D. C. & Anderson, D. L.) 71–90
- 485 (2005). doi:10.1130/0-8137-2388-4.71
- 486 95. Bohrson, W. a. *et al.* Prolonged history of silicic peralkaline volcanism in the eastern Pacific Ocean.
- 487 *J. Geophys. Res.* **101**, 11457 (1996).
- 488 96. Kroenke, L. W. Plate tectonic development of the western and southwestern Pacific: mesozoic to
- 489 the present. in *The origin and evolution of Pacific Island biotas, New Guinea to Eastern Polynesia:*
- 490 *patterns and processes* (eds. Keast, A. & Miller, S.) 19–34 (1996).
- 491 97. Ollier, C. D. Geomorphology of South Atlantic volcanic islands. Part I: the Tristan da Cunha group.
- 492 *Zeitschrift fur Geomorphol.* **28**, 367–382 (1984).
- 493 98. Kocher, T. D. *et al.* Dynamics of mitochondrial DNA evolution in animals: amplification and
- 494 sequencing with conserved primers. *Proc. Natl. Acad. Sci. U. S. A.* **86**, 6196–6200 (1989).
- 495 99. Dietzen, C., Witt, H.-H. & Wink, M. The phylogeographic differentiation of the European robin
- 496 *Erithacus rubecula* on the Canary Islands revealed by mitochondrial DNA sequence data and
- 497 morphometrics: evidence for a new robin taxon on Gran Canaria? *Avian Sci.* **3**, 115–132 (2003).
- 498 100. Edwards, S. V., Arctander, P. & Wilson, A. C. Mitochondrial resolution of a deep branch in the
- 499 genealogical tree for perching birds. *Proc. Biol. Sci.* **243**, 99–107 (1991).
- 500 101. Helm-Bychowski, K. & Cracraft, J. Recovering phylogenetic signal from DNA sequences:
- 501 relationships within the corvine assemblage (class aves) as inferred from complete sequences of
- 502 the mitochondrial DNA cytochrome-b gene. *Mol. Biol. Evol.* **10**, 1196–1214 (1993).
- 503 102. Warren, B. H., Bermingham, E., Bowie, R. C. K., Prys-Jones, R. P. & Thébaud, C. Molecular
- 504 phylogeography reveals island colonization history and diversification of western Indian Ocean
- 505 sunbirds (Nectarinia: Nectariniidae). *Mol. Phylogenet. Evol.* **29**, 67–85 (2003).
- 506 103. Farrington, H. L., Lawson, L. P., Clark, C. M. & Petren, K. The evolutionary history of Darwin's
- 507 finches: speciation, gene flow, and introgression in a fragmented landscape. *Evolution* (2014).

- 508 doi:10.1111/evo.12484
- 509 104. Warren, B. H. *et al.* Hybridization and barriers to gene flow in an island bird radiation. *Evolution*
510 (N. Y). **66**, 1490–1505 (2012).
- 511 105. Warren, B. H., Bermingham, E., Prys-Jones, R. P. & Thebaud, C. Tracking island colonization history
512 and phenotypic shifts in Indian Ocean bulbuls (Hypsipetes: Pycnonotidae). *Biol. J. Linn. Soc.* **85**,
513 271–287 (2005).
- 514 106. Andersen, M. J., Hosner, P. A., Filardi, C. E. & Moyle, R. G. Phylogeny of the monarch flycatchers
515 reveals extensive paraphyly and novel relationships within a major Australo-Pacific radiation.
516 *Mol. Phylogenet. Evol.* **83**, 118–136 (2015).
- 517 107. Sari, E. H. R. & Parker, P. G. Understanding the colonization history of the Galápagos flycatcher
518 (*Myiarchus magnirostris*). *Mol. Phylogenet. Evol.* **63**, 244–54 (2012).
- 519 108. Chaves, J. A., Parker, P. G. & Smith, T. B. Origin and population history of a recent colonizer, the
520 yellow warbler in Galápagos and Cocos Islands. *J. Evol. Biol.* **25**, 509–21 (2012).
- 521 109. Martínez-Gómez, J. E., Barber, B. R. & Peterson, a T. Phylogenetic position and generic placement
522 of the Socorro wren (*Thryomanes Sissonii*). *Au* **122**, 50–56 (2005).
- 523 110. Warren, B. H., Bermingham, E., Prys-Jones, R. P. & Thébaud, C. Immigration, species radiation and
524 extinction in a highly diverse songbird lineage: White-eyes on Indian Ocean islands. *Mol. Ecol.* **15**,
525 3769–3786 (2006).
- 526 111. McGuire, J. A. *et al.* Molecular phylogenetics and the diversification of hummingbirds. *Curr. Biol.*
527 **24**, 910–916 (2014).
- 528 112. Derryberry, E. P. *et al.* Lineage diversification and morphological evolution in a large-scale
529 continental radiation: The neotropical ovenbirds and woodcreepers (aves: furnariidae). *Evolution*
530 (N. Y). **65**, 2973–2986 (2011).
- 531 113. Jønsson, K. A. *et al.* A supermatrix phylogeny of corvid passerine birds (Aves: Corvides). *Mol.*
532 *Phylogenet. Evol.* **94**, 87–94 (2016).
- 533 114. Scofield, R. P. *et al.* The origin and phylogenetic relationships of the New Zealand ravens. *Mol.*
534 *Phylogenet. Evol.* **106**, 136–143 (2017).
- 535 115. Cibois, A., Thibault, J. C., Bonillo, C., Filardi, C. E. & Pasquet, E. Phylogeny and biogeography of the
536 imperial pigeons (Aves: Columbidae) in the Pacific Ocean. *Mol. Phylogenet. Evol.* **110**, 19–26
537 (2017).
- 538 116. Friis, G., Aleixandre, P., Rodríguez-Estrella, R., Navarro-Sigüenza, A. G. & Milá, B. Rapid postglacial
539 diversification and long-term stasis within the songbird genus Junco: phylogeographic and
540 phylogenomic evidence. *Mol. Ecol.* **25**, 6175–6195 (2016).
- 541 117. Marki, P. Z. *et al.* Supermatrix phylogeny and biogeography of the Australasian Meliphagides
542 radiation (Aves: Passeriformes). *Mol. Phylogenet. Evol.* **107**, 516–529 (2017).
- 543 118. Fuchs, J. *et al.* Long-distance dispersal and inter-island colonization across the western Malagasy
544 Region explain diversification in brush-warblers (Passeriformes: Nesillas). *Biol. J. Linn. Soc.* **119**,
545 873–889 (2016).
- 546 119. Cibois, A. *et al.* Phylogeny and biogeography of the fruit doves (Aves: Columbidae). *Mol.*
547 *Phylogenet. Evol.* **70**, 442–453 (2014).
- 548 120. Carmi, O., Witt, C. C., Jaramillo, A. & Dumbacher, J. P. Phylogeography of the Vermilion Flycatcher
549 species complex: Multiple speciation events, shifts in migratory behavior, and an apparent
550 extinction of a Galápagos-endemic bird species. *Mol. Phylogenet. Evol.* **102**, 152–173 (2016).
- 551 121. Cornetti, L. *et al.* The genome of the ‘great speciator’ provides insights into bird diversification.
552 *Genome Biol. Evol.* **7**, 2680–2691 (2015).
- 553



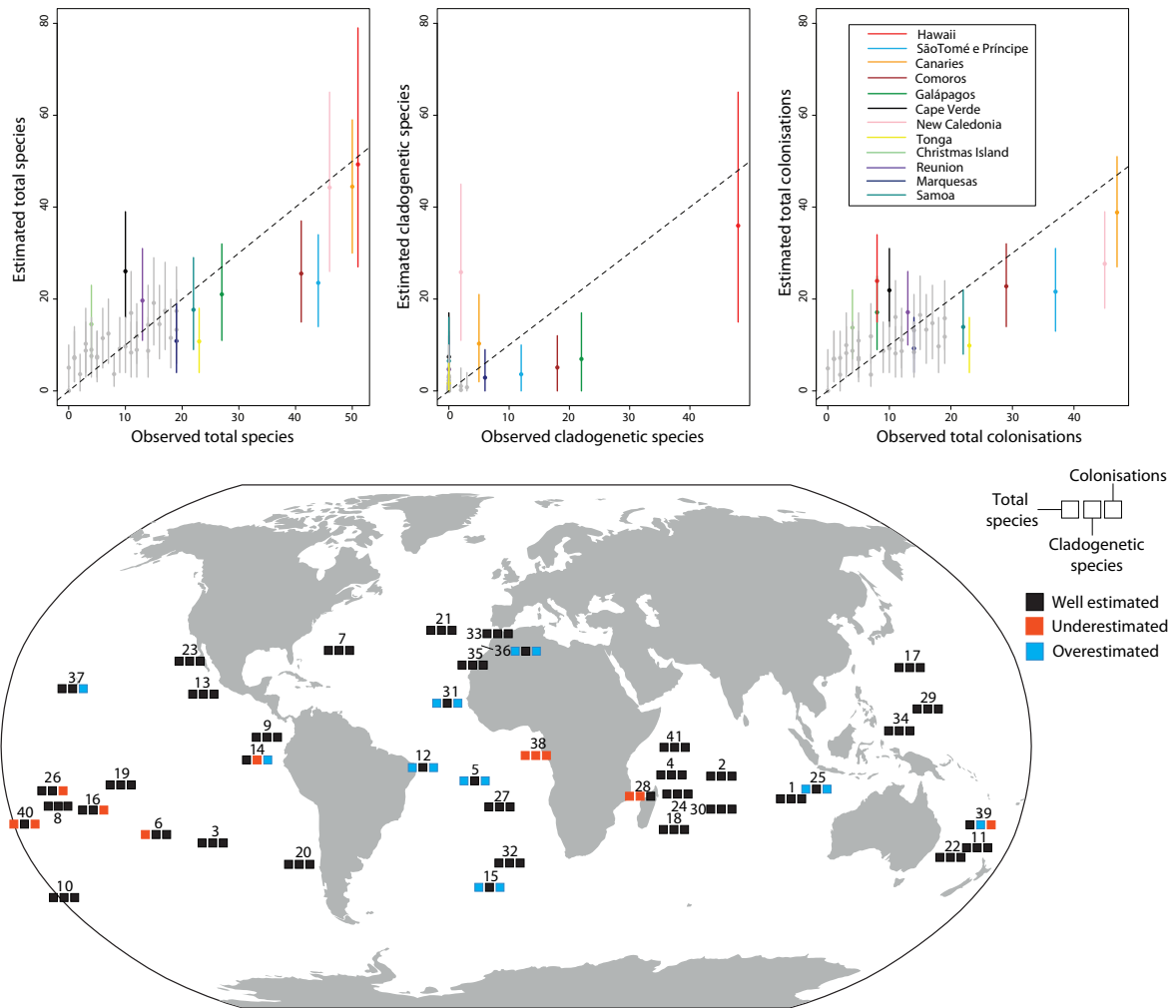
554

555 **Figure 1 – Archipelago and island bird colonisation time data.** Archipelagos
556 included in the dataset, showing the total number of species belonging to our focal
557 group (both extinct and extant) found in each archipelago (at the time of human
558 arrival). Numbers on map correspond to numbers to the left of the archipelago name.
559 Numbers to the right of the archipelago name: number of species from our focal
560 assemblage on the archipelago | percentage of species sampled in the phylogenetic
561 trees. Even species not sampled in the trees are accounted for by including them as
562 missing species that could have colonised any time since emergence of the archipelago.
563 Colonisation times plot: grey horizontal lines – archipelago ages (Extended Data Table
564 1). Violin plots (blue) show the kernel density of the distribution of times of
565 colonisation of bird species in each archipelago, obtained from the time-calibrated
566 phylogenetic trees. Thick black line inside violin - interquartile distance; thin black line -
567 95% CI; black dot - median. Archipelagos with no violin plot or dots are cases for which
568 no species of our focal assemblage were present at the time of human arrival, or none
569 were sampled using molecular data. Birds from left to right: Seychelles sunbird,
570 Seychelles magpie robin, silvereye, Príncipe thrush, laurel pigeon, dodo (extinct),
571 Mauritius fody, red-moustached fruit dove (extinct), Galápagos warbler, Norfolk kaka
572 (extinct). Bird photos and illustrations used with permission from: Cláudia Baeta, Pedro
573 Cascão, Martijn Hammers, Dubi Shapiro and Juan Varela. There are no *in-situ* radiations
574 in the Mascarenes (Mauritius, Reunion and Rodrigues) because we treat the islands as
575 separate entities (but see sensitivity analyses).



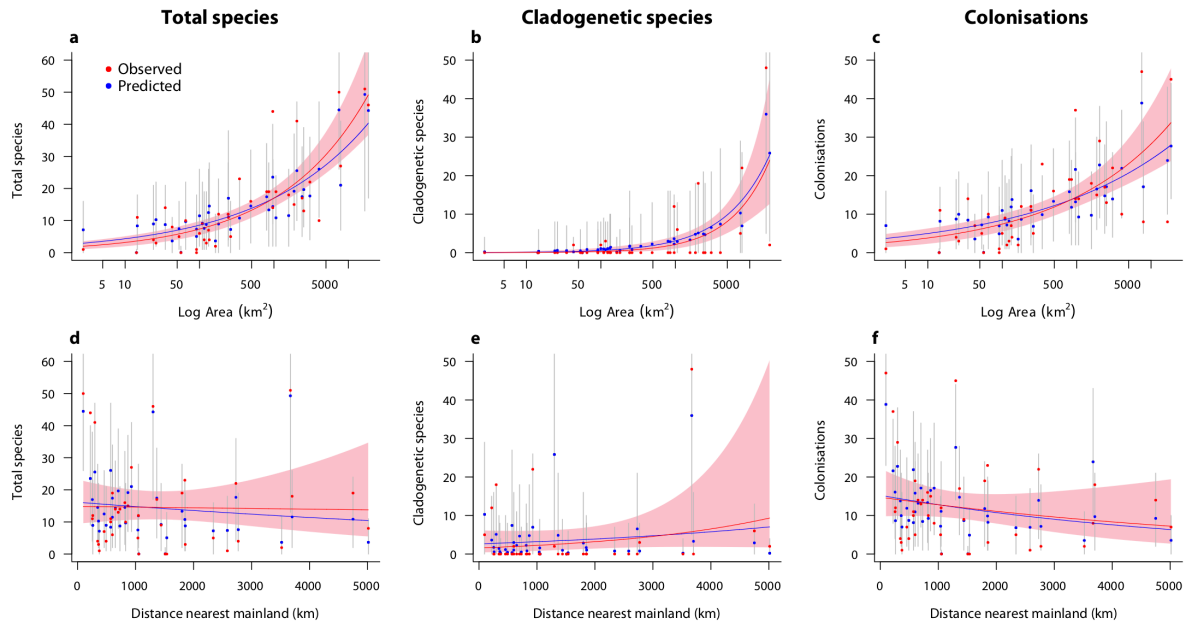
576
 577
 578
 579
 580
 581
 582
 583
 584
 585
 586
 587

Figure 2 - Estimated relationships between island area and isolation and local island biogeography parameters. Isolation measured as distance to the nearest mainland (D_m). Assuming the maximum likelihood global hyperparameters of the best models (equations describing the relationships given in Supplementary Information 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. Extinction rate depends on area. Colonisation rate and anagenesis rate depend on D_m . See also Extended Data Fig. 1 for the relationship of cladogenesis with both area and distance under the M19 model.



588
 589
 590
 591
 592
 593
 594
 595
 596
 597

Figure 3 – Goodness of fit of the preferred model (M19). Top 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. Dashed line is $y=x$. Selected archipelagos are highlighted in colour. 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) for each archipelago. Numbers indicating archipelagos on the map match those in Fig. 1.



598

599 **Figure 4 – Observed and simulated island diversity-area and island diversity-**
 600 **distance relationships.** Grey vertical lines show the 95% confidence intervals across
 601 1,000 datasets simulated for each of the 41 archipelagos assuming the M19 model. Blue
 602 points: mean values of the simulations; blue line – fitted line for the simulated data; red
 603 points – observed values in the empirical data; red line – fitted line for the empirical
 604 data; red shaded area is the 95% confidence interval of the predicted relationship for
 605 the empirical data.
 606

607 **Methods**

608

609 **Archipelago selection**

610

611 We focus on oceanic islands, i.e. volcanic islands that have never been connected to any
612 other landmass in the past. We also include the Granitic Inner Seychelles, even though
613 these islands have a continental origin, because they have been separated from other
614 landmasses for a very long period of time (64 million years (Ma)³³) and can be
615 considered quasi-oceanic, as all extant avian species originated in much more recent
616 times. The 41 archipelagos chosen are located in the Atlantic, Indian and Pacific oceans,
617 with latitudes between 45° North and South. Islands within these archipelagos are
618 separated by a maximum of 150 km. The sole exceptions are the Azores and Hawaii, two
619 very isolated systems where the distances between some islands exceed this value.

620

621 **Physical and geological data**

622

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

633 Island palaeo-areas and past archipelago configurations have been shown to be
634 better predictors of endemic insular diversity than contemporary area^{14,34}. In contrast,
635 island total native and non-endemic richness is better predicted by present island
636 characteristics^{14,34}. With insufficient data on island ontogeny being available (i.e.
637 describing empirical area trajectories from island birth to present) we therefore
638 analysed contemporary area and isolation as currently the most appropriate units for
639 our dataset.

640 We conducted an extensive survey of the literature and consulted geologists to
641 obtain archipelago geological ages (Extended Data Table 1), treating the age of the
642 oldest currently emerged island as an upper bound for colonisation. Islands may have
643 been submerged and emerged multiple times and we consider the age of the last known
644 emergence. For the Aldabra Group we used an age older than the published estimate.
645 The current estimated age of re-emergence of Aldabra is 0.125 Ma³⁵, but nine out of 12
646 Aldabra colonisation events in our dataset are older, suggesting the archipelago was not
647 fully submerged prior to this and may have been available for colonisation for a longer
648 period. Therefore, for Aldabra we used an older upper bound of 1 Ma for colonisation,
649 although we acknowledge that the mitochondrial markers used for dating may not
650 provide sufficient resolution at the shallow temporal scale of the published age. For
651 Hawaii, the colonisation times we obtained for more than half of the colonisation events
652 were older than the age of the current high islands that is often used as a maximum age
653 for colonisation (~5 Ma). Therefore, instead of this age, we used the much older
654 estimate of 29.8 Ma of the Kure Atoll³⁶ to account for currently submerged or very low-
655 lying Hawaiian islands that could have received colonists in the past. For Bermuda and
656 Marianas, we could not find age estimates in the literature, and we therefore consulted
657 geologists to obtain these (P. Hearty, R. Stern and M. Reagan, pers. comm., Extended
658 Data Table 1).

659

660 **Island avifaunas**

661

662 Our sampling focused on native resident terrestrial birds and we considered only birds
663 that colonise by chance events (e.g. hurricanes, raft). We thus excluded marine and
664 migratory species, because they are capable of actively colonising an island at a much
665 higher rate. We focused on songbird-like and pigeon-like birds, which constitute the
666 majority of terrestrial (land-dwelling) birds on islands. Following a precedent set by
667 previous work^{9,25,30}, we included only species from the same trophic level (in the spirit
668 of MacArthur and Wilson's model): we excluded aquatic birds, birds of prey, rails (many
669 are flightless or semi-aquatic) and nightjars (nocturnal). We also excluded introduced
670 and vagrant species. Including species such as rails and owls (which are components of
671 many island avifaunas) would have led to a higher estimate of the product of
672 colonisation rate and mainland pool size due to a larger mainland pool, and potentially

673 to higher estimated rates of anagenesis (many owl or rail species are island endemics
674 with no close relatives on the islands).

675 For the focal avian groups, we compiled complete taxon lists for each of the 41
676 archipelagos based on recent checklists from Avibase (<http://avibase.bsc-eoc.org>),
677 which we cross-checked with the online version of the *Handbook of the Birds of the*
678 *World* (HBW³⁷). We followed HBW's nomenclature and species assignments, except for
679 12 cases where our phylogenetic data disagree with HBW's scheme (noted in the
680 column 'Taxonomy' of Supplementary Data 1). For example, in eleven cases
681 phylogenetic trees support raising endemic island subspecies to species status (we
682 sampled multiple samples per island taxon and outgroup, and the island individuals
683 form a reciprocally monophyletic well-supported clade), and for these taxa we decided
684 it was more appropriate to use a phylogenetic species concept so as not to
685 underestimate endemism and rates of speciation (Supplementary Data 1). We re-ran
686 DAISIE analyses using HBW's classification and found that the maximum likelihood
687 parameters are very similar and thus we report only the results using the scheme based
688 on the phylogenies produced for this study.

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

693

694

695 **Extinct species**

696 We do not count extinctions with anthropogenic causes as impacting the natural
697 background rate of extinction. Therefore, we explicitly include species where there is
698 strong evidence that they have been extirpated by humans. We treat taxa extirpated on
699 an archipelago by humans as though they had survived in that archipelago until the
700 present following the approach of ref³⁸.

701 We identified anthropogenic extinctions based on published data³⁹⁻⁴⁶ and
702 personal comments (Josep Antoni Alcover and Juan Carlos Rando on unpublished
703 Macaronesian taxa; Ferran Sayol and Søren Faurby). We include the species present on
704 the islands that belong to our archipelago definition as in Supplementary Data 2. We
705 excluded largely hypothetical accounts or pre-Holocene fossils that greatly pre-date

706 human arrival. Our dataset accounts for 153 taxa that were present upon first human
707 contact and have gone extinct since, probably because of human activities including
708 human introduction of invasive species. To our knowledge 71 of these taxa have
709 previously been sequenced using ancient DNA or belong to clades present in our trees,
710 and we were thus able to include them in the phylogenetic analyses as regular data (n =
711 54), or as missing species by adding them as unsampled species to a designated clade (n
712 = 17). For the remaining 82 extinct taxa, sequences were not available and we were
713 unable to obtain samples and to allocate them to clades. We assume that these taxa
714 represent extinct independent colonisations and we included them in the analyses using
715 the “Endemic_MaxAge” and “Non_endemic_MaxAge” options in DAISIE, which assume
716 that they have colonised at any given time since the birth of the archipelago (but before
717 any *in situ* cladogenesis event). As an example, our dataset includes the 27 species of
718 Hawaiian birds belonging to our focal group that are known to have gone extinct since
719 human colonisation. Eight of these species were included using DNA data, 17 were
720 added as missing species to their clades (14 honeycreepers and 3 *Myadestes*) and two
721 were added using the Endemic_MaxAge option in DAISIE (*Corvus impluviatus* and
722 *Corvus viriosus*).

723

724 **Sequence data: GenBank**

725

726 We conducted an extensive search of GenBank for available DNA sequences from the
727 596 island bird taxa fitting our sampling criteria and from multiple outgroup taxa, using
728 software Geneious 11⁴⁷. The molecular markers chosen varied from species to species,
729 depending on which marker was typically sequenced for the taxon in question, the
730 commonest being cytochrome b (*cyt-b*). In total, we downloaded 3155 sequences from
731 GenBank. For some taxa, sequences from both archipelago and close relatives from
732 outside the archipelago were already available from detailed phylogenetic or
733 phylogeographic analyses. In some cases, a target species had been sampled, but only
734 from populations outside the archipelago. In other cases, the species on the archipelago
735 had been sampled, but the sampling of the relatives outside of the archipelago was
736 lacking or only from distant regions, which meant a suitable outgroup was not available
737 on GenBank. Finally, for some species there were no previous sequences available on
738 GenBank. GenBank accession numbers and geographical origin for the downloaded

739 sequences are provided on the maximum clade credibility trees (deposited in
740 Mendeley).

741 **Sequence data: new samples**

742

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

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

768 Polymerase chain reactions (PCR) were set up in 25 μ L total volumes including 5
769 μ L of buffer BioLone MyTaq, 1 μ L (10 mM) of each primer, and 0.12 μ L MyTaq
770 polymerase. PCRs were performed with the following thermocycler conditions: initial
771 denaturation at 95° C for 1 min followed by 35 cycles of denaturation at 95° C for 20 s,

772 with an annealing temperature of 48° C for 20 s, and extension at 72° C for 15 s min and
773 a final extension at 72° C for 10 min. Amplified products were purified using
774 Exonuclease I and Antarctic Phosphatase, and sequenced at the University of Potsdam
775 (Unit of Evolutionary Biology/Systematic Zoology) on an ABI PRISM 3130xl sequencer
776 (Applied Biosystems) using the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied
777 Biosystems). We used Geneious 11 to edit chromatograms and align sequences.

778

779

780 **Phylogenetic analyses**

781

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

795 For the new dating analyses conducted for this study, we created 80 separate
796 alignments for different groups using a combination of sequences from Genbank (n =
797 3155) and new sequences (n = 252) produced for this study. In some cases, we obtained
798 alignments directly from authors of previous studies and these are credited in Extended
799 Data Table 3. Phylogenetic divergence dating analyses were performed in BEAST 2⁴⁸.
800 For each alignment we performed substitution model selection in jModeltest⁴⁹ using the
801 Bayesian information criterion. We used rates of molecular evolution for avian
802 mitochondrial sequences, which have been shown to evolve in a clock-like fashion at an
803 average rate of ~2% per Ma⁵⁰. Molecular rate calibrations can be problematic for
804 ancient clades, due to high levels of heterotachy in birds⁵¹. In addition, mitochondrial

805 DNA saturates after about 10 to 20 million years, and genetic distances of more than
806 20% may provide limited information regarding dating⁵². Therefore, we only used
807 molecular rate dating to extract node ages for branching events at the tips of the trees,
808 at the species or population level (oldest colonisation time in our dataset is 15.3 Ma, but
809 most are much younger). Rates of evolution were obtained from the literature and
810 varied between different markers and taxonomic group (Supplementary Information
811 Table 4). We applied the avian mitochondrial rates estimated from *cyt-b* by Weir and
812 Schluter⁵⁰ (but see ‘Sensitivity to alternative branching times and tree topologies’
813 section for different rates).

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

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

825

826 **Colonisation and branching times**

827 The nodes selected in the dated trees for estimates of colonisation and branching times
828 are given for each taxon in Supplementary Data 1. Our node selection approach was as
829 follows. For cases in which archipelago individuals formed a monophyletic clade
830 consisting exclusively of archipelago individuals, we used the stem age of this clade as
831 colonisation time. For cases in which only one individual of the archipelago was
832 sampled, we used the length of the tip leading to that individual, which is equivalent to
833 the stem age. For cases in which the archipelago individuals were embedded in a clade
834 containing mainland individuals of the same species, i.e. paraphyly or polyphyly; we
835 assumed that this is due to incomplete lineage sorting of the insular and mainland
836 lineages, and we therefore used the MRCA of the archipelago individuals, or the crown
837 node when the MRCA coincides with the crown. For these later cases using the stem

838 would most likely have been an overestimation of the colonisation time, as we assume
839 that colonisation happens from the mainland to the archipelago. For these later cases
840 we applied the ages using the “MaxAge” option in DAISIE, which integrates over the
841 possible colonisation times between the present and the upper bound. A robustness test
842 of our results to node choice is given in section “Sensitivity analyses: alternative
843 branching times and tree topologies”.

844 For a total of 19 endemic taxa we could not obtain sequences, but we could
845 allocate them to a specific island clade (e.g. Hawaiian honeycreepers and solitaires).
846 These were added as missing species to that clade. For 96 non-endemic taxa we could
847 not obtain sequences of individuals from the archipelago, but we could obtain
848 sequences from the same species from different regions. For these cases we used the
849 crown or the stem age of the species as an upper bound for the age of the colonisation
850 event, using the “Non_endemic_MaxAge” option in DAISIE. Finally, for 124 taxa (20.8 %)
851 no sequences of individuals from the archipelago were available on GenBank and we
852 were not able to obtain samples for sequencing from the species or from close relatives.
853 We assumed these cases constituted independent colonisations that could have taken
854 place any time since the origin of the archipelago and the present, and applied the
855 “Non_endemic_MaxAge” and “Endemic_MaxAge” options in DAISIE with a maximum age
856 equal to the archipelago age. DAISIE makes use of this information⁵³.

857

858 **Global dataset characteristics**

859

860 Data points from taxa of the same archipelago were assembled into 41 archipelago-
861 specific datasets. These 41 datasets were in turn were assembled into a single dataset –
862 hereafter the “global dataset” - which was analysed with DAISIE (DAISIE R object
863 deposited in Mendeley). The global dataset (information Supplementary Data 1) has a
864 total of 596 taxa (independent colonisation events plus species within radiations),
865 covering 491 species from 203 different genera and eight orders. All taxa were included
866 in the analyses: those which we sampled in phylogenies, but also those for which
867 sequences or phylogenies could not be obtained and which were included following the
868 approaches described in the *colonisation and branching time* methods. A summary of
869 diversity and sampling per archipelago is given in Extended Data Table 1.

870

871 **Sampling completeness**

872

873 In total, we produced new sequences from 252 new individuals, comprising 90 different
874 species from 45 different genera, covering an additional 110 colonisation events that
875 had never before been sampled (i.e. populations from islands where the species had not
876 been sampled before). For at least 12 of these 90 species, we found no previous
877 sequences on Genbank, including island endemics from Comoros, Galápagos, Rodrigues
878 and São Tomé (Supplementary Information Table 5). The new sequences from 252
879 individuals increase the molecular sampling for extant colonisation events from 60%
880 (223/373) to 89% (332/373). If we include historically extinct colonisations, we
881 increased the molecular sampling from the existing 54% (269/502) of colonisation
882 events to 75% (379/502). We also substantially increased molecular sampling of
883 continental relatives, adding 78 new individuals from the continent or islands
884 surrounding our archipelagos, covering 43 different species. The percentage of taxa
885 sampled in phylogenies varied widely between archipelagos (Extended Data Table 1
886 and Fig. 1). For eight archipelagos (Bermuda, Fernando de Noronha, Pitcairn, Rapa Nui,
887 Rodrigues, Saint Helena, Society Islands and Tonga) less than 50% of the species were
888 sampled in phylogenies, and thus the majority of the species for these island groups
889 were added with maximum ages and endemism status. For 13 archipelagos, which
890 accounted for more than a third of the total species, over 90% of the species were
891 sampled in phylogenies.

892

893 **DAISIE**

894

895 We used the method DAISIE⁹ (Dynamic Assembly of Islands through Speciation,
896 Immigration, and Extinction) to estimate rates of species accumulation (colonisation,
897 speciation and extinction) on the archipelagos. The model assumes that after the origin
898 of an island, species can colonise from a mainland pool. Once a species has colonised, it
899 may remain similar to its mainland ancestor (non-endemic species), become endemic
900 through anagenetic speciation (new endemic species is formed without lineage splitting
901 on the island), split into new species via cladogenetic speciation and/or go extinct. A
902 carrying capacity (i.e. maximum number of species each colonist lineage can attain) is

903 implemented, such that rates of cladogenesis and colonisation decline with increasing
904 number of species in the colonising clade.

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

914 In its parameterization of extinction, DAISIE accounts for the fact that there may
915 have been several lineages that were present on the insular system in the past but
916 which went completely extinct due to natural causes, leaving no extant descendants.
917 Simulations have shown that the rate of natural extinction is usually well estimated in
918 DAISIE (Methods section *Measuring precision and accuracy* and ref. ⁵³). Studies on
919 phylogenies of single clades suggest that phylogenetic data on only extant species
920 provide less information on extinction than on speciation (or rather diversification
921 rates⁵⁴). However, there is information-content in such data⁵⁵, especially when
922 diversification dynamics are diversity-dependent⁵⁶. Moreover, here we use colonisation
923 times in addition to phylogenetic branching times to estimate extinction rates, and we
924 are estimating hyperparameters that theory suggests correlate with extinction (i.e.
925 area). Finally, we use data from many independent colonisations, which increases the
926 power of our statistical method considerably, and decreases the bias, as maximum
927 likelihood is known to asymptotically provide unbiased estimates.

928

929 **Estimating global hyperparameters**

930

931 Our aim is to examine the dependencies of the parameters that govern species assembly
932 (colonisation, extinction, cladogenesis, anagenesis (CES rates), and carrying capacity)
933 on the features of archipelagos (area, isolation). We developed a new method to
934 estimate global hyperparameters that control the relationship between two key
935 archipelago features (area and isolation) and archipelago-specific (local) CES rates. One

936 can estimate directly from the global dataset the shape of the relationship between
937 isolation and colonisation rate that maximizes the likelihood for the entire dataset.

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

956 In addition to the *a priori* models, we considered a set of *post hoc* models with
957 alternative shapes of relationships. We fitted two types of *post hoc* models: power
958 models and sigmoid models (Supplementary Information Table 1). In the *post hoc*
959 power models we modelled all parameters as in the *a priori* models, except for
960 cladogenesis: we allowed cladogenesis to be dependent both on area and distance. The
961 reason for this is that we found that the predicted number of cladogenetic species under
962 the *a priori* models were not as high as observed, so we examined whether including a
963 positive effect of distance would improve the fit. We described the relationship between
964 area, distance and cladogenesis using different functions – one model where there is an
965 additive effect of area and distance (M15); and three models (M16, M17, M18) where
966 the effect of area and distance is interactive. In addition, we fitted a model identical to
967 M16 but with one parameter less (M19). The reason for this was that this parameter (y)
968 was being estimated as zero in M16

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

979 In total we fitted 28 candidate models (14 *a priori*, 14 *post hoc*) to the global
980 dataset using maximum likelihood (ML). We fitted each model using 20 initial sets of
981 random starting parameters to reduce the risk of being trapped in local likelihood
982 suboptima. We used the age of each archipelago (Extended Data Table 1) as the
983 maximum age for colonisation. We assumed a global mainland species pool M of 1000
984 species. The product of M and the intrinsic rate of colonisation (γ_0) is constant as long
985 as M is large enough (larger than the number of island species), and thus the chosen
986 value of M does not affect the results.

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

996 An alternative approach to estimating hyperparameters would be to calculate CES
997 rates and their uncertainty independently for each archipelago and to then conduct a
998 meta-analysis of the resulting data, including archipelago area and isolation as
999 predictors. However, errors in parameter estimates will vary, particularly because some
1000 archipelagos have small sample sizes (only a few extant colonisation events, or none at
1001 all, e.g. Chagos) and are thus much less informative about underlying process⁵³. Thus,

1002 maximizing the likelihood of all data sets together by estimating the hyperparameters
1003 (which is precisely our aim) is preferable. For completeness, we present CES rates
1004 estimated independently for each archipelago in Supplementary Information Table 6,
1005 excluding archipelagos with fewer than six species and for which we sampled less than
1006 60% of the species in the phylogenies. However, as argued above we do not advocate
1007 using these parameter estimates for further analyses because the number of taxa for
1008 some of these archipelagos is still low and by excluding archipelagos with fewer than six
1009 taxa we cannot capture the lower part of the relationship between area/isolation and
1010 CES rates.

1011 All DAISIE analyses were run using parallel computation on the high-performance
1012 computer clusters of the University of Groningen (Peregrine cluster) and the Museum
1013 für Naturkunde Berlin. The new version of the R package DAISIE is available on Github.

1014

1015

1016 **Randomization analysis**

1017

1018 We conducted a randomization analysis to evaluate whether there is significant signal
1019 of a relationship between area and distance and local CES rates in our global dataset.
1020 We produced 1,000 datasets with the same phylogenetic data and archipelago ages as
1021 the global dataset, but randomly reshuffled archipelago area and D_m in each dataset. We
1022 then fitted the best *post hoc* model to each of these 1,000 randomized datasets. If the
1023 maximum likelihood estimates of exponent hyperparameters (i.e. the strength of the
1024 relationship) in the randomized datasets were non-zero this would indicate that the
1025 method is finding evidence for a relationship even if there is none. If, on the other hand,
1026 non-zero hyperparameters are estimated in the real data but not in the randomized
1027 datasets, this would mean that there is information in the data regarding the putative
1028 relationships.

1029 The randomization analysis showed that in global datasets with reshuffled areas
1030 and distances the exponent hyperparameters are estimated as zero in most cases,
1031 whereas in the empirical global dataset they are not (Extended Data Fig. 4).

1032

1033 **A posteriori simulations**

1034

1035 We simulated 1,000 phylogenetic global datasets (41 archipelagos each) with the ML
1036 hyperparameters of the best *a priori* (M14) and *post hoc* models (M19). We first
1037 calculated the local CES rates for each archipelago based on their area and isolation and
1038 the hyperparameters for the model, and then used these CES rates as the parameters for
1039 the simulations using the DAISIE R package. The simulated data were used to measure
1040 bias and accuracy of the method, goodness of fit and the ability of our method to recover
1041 observed island biogeographic diversity patterns (see below).

1042

1043 **Measuring precision and accuracy of method**

1044

1045 DAISIE estimates CES rates with high precision and little bias^{9,53}. We conducted
1046 parametric bootstrap analyses to assess whether the ability to estimate
1047 hyperparameters from global datasets is also good (Extended Data Fig. 2 and 3), and to
1048 obtain confidence intervals on parameter estimates (Extended Data Table 5). We used
1049 DAISIE to estimate hyperparameters from the M14 and M19 simulated datasets (1,000
1050 replicates each). We measured precision and accuracy by comparing the distribution of
1051 parameters estimated from the 1,000 simulated data set with the real parameters used
1052 to simulate the same datasets. To check whether ML optimisations of the simulated
1053 global datasets converge to the same point in parameter space, we first performed a test
1054 on a subset of the simulated data. We ran optimisations with 10 random sets of initial
1055 starting values for each of 10 simulated datasets. All optimisations converged to the
1056 same likelihood and a very similar hyperparameter set; therefore, we are confident we
1057 found the global optimum for each simulated global dataset, even for models with many
1058 parameters.

1059

1060 **Measuring goodness of fit**

1061

1062 We measured how well the preferred models fitted the data using different approaches.
1063 First, we examined whether our models successfully reproduce the diversity patterns
1064 found on individual archipelagos. We calculated the total number of species,
1065 cladogenetic species and independent colonisations in each archipelago for each of the
1066 1,000 simulated datasets. We then plotted these metrics versus the observed values in
1067 the empirical data (Fig. 3). Our preferred models have a slight tendency to overpredict

1068 species richness when there are a few species and underpredict it when there are many.
1069 We do not have a clear explanation for this. This slight deviation does not seem to be
1070 due to an additional area- or distance-dependence, so explanation should be sought in
1071 other factors that we did not model. We note that the fact that all three plots show this
1072 tendency rather than just one is to be expected because the three metrics of species
1073 richness are not entirely independent, with total species richness being the sum of the
1074 other two.

1075 Second, we examined whether the models successfully predict the empirical
1076 relationships between area, distance and diversity metrics (total species, cladogenetic
1077 species, and number of independent colonisations). We fitted generalised linear models
1078 (GLM) for each diversity metric, with quasipoisson family errors and log area (or
1079 distance) as predictors. We then repeated this across 1000 independent sets of
1080 simulated data for the 41 archipelagos and compared the distribution of slopes and
1081 intercepts for archipelago area and archipelago isolation to the equivalent estimates for
1082 the empirical data (Fig. 4).

1083 Third, we estimated the pseudo- R^2 of the best model (M19) as a measure of the
1084 model's explanatory power. We simulated two independent sets of 10,000 global
1085 datasets under M19 model (Set 1 and Set 2). We calculated the mean total number of
1086 species, number of cladogenetic species and colonisations for each archipelago across
1087 all datasets from Set 1. For each diversity metric we calculated a pseudo- R^2 (pseudo- R^2 -
1088 observed) where the total sum of squares was obtained from the empirical data and the
1089 residual sum of squares was obtained as the difference between empirical values and
1090 expected values (i.e. the simulation means). As the model is inherently stochastic, even
1091 if the model was an accurate and complete reflection of the underlying processes then
1092 the pseudo- R^2 would tend to be < 1 . To estimate the distribution of pseudo- R^2 expected
1093 under the model we treated the set 2 simulations as data and estimated the pseudo- R^2
1094 for each (pseudo- R^2 -simulated). We then calculated the ratio of the pseudo- R^2 -observed
1095 values over the 10,000 pseudo- R^2 -simulated values. A ratio approaching 1 would
1096 indicate that the model is explaining the observed data as well as the average dataset
1097 simulated under this process (Extended Data Fig. 5).

1098

1099 **Sensitivity to alternative divergence times and tree topology**

1100

1101 Despite having sampled several new individuals from islands worldwide, given the wide
1102 geographical scale of our study we still rely on sequence data for thousands of
1103 individuals which has been submitted to Genbank over the years. Whenever multi-loci
1104 analyses including our focal taxa were available we used them, but these are rare
1105 (Extended Data Table 4). Therefore, the majority of our phylogenies are based on a
1106 small number of genes, and most on a single gene, *cyt-b*, which is the most widely
1107 sequenced mitochondrial marker in birds. Although some studies on island birds have
1108 shown that colonisation times derived from mitochondrial trees often do not differ
1109 much from those obtained using multiple loci (Stervander et al, MORE REFS?), it is
1110 possible that for some cases the scaling and topologies of the trees might have been
1111 more accurate had we used multiple loci⁵⁷. This is particularly relevant for recent island
1112 colonists, given incomplete lineage sorting⁵⁸. An additional shortcoming of relying on
1113 published sequence data is that many of our DNA alignments often have substantial
1114 sections with missing data (e.g. because only one small section of the gene could be
1115 sequenced and was uploaded to Genbank), which has been shown to lead to biases in
1116 branch lengths and topology⁵⁹. While future studies using phylogenomic approaches
1117 may address these issues, obtaining tissue samples for all these taxa will remain an
1118 obstacle for a long time.

1119 Although DAISIE does not directly use topological information (only divergence
1120 times are used), it is possible that the true **topology** for a clade may differ from that of
1121 the gene tree we have estimated and this could have an impact on our results by a)
1122 affecting colonisation and branching times (addressed in the paragraph below); or b) by
1123 altering the number of colonisation events. For instance, species that appear to have
1124 colonised an archipelago only once may have colonised multiple times. As with any
1125 phylogenetic study, we cannot rule out this possibility, but we assume that re-
1126 colonisation of any of the archipelagos in our dataset by the same species is very rare, as
1127 these are all oceanic and isolated. For archipelago lineages with cladogenesis (26 out of
1128 502 lineages), alternative topologies could include non-monophyly of island radiations,
1129 with the corollary being that they would be the results of multiple colonisations.
1130 However, non-monophyly of these well-studied and highly isolated insular radiations is
1131 highly improbable.

1132 Regarding **scaling of divergence times**, we assessed how uncertainty in our
1133 estimated node ages could influence our results by running an analysis of 100 datasets.

1134 For each dataset we sampled the node ages (i.e. colonisation and branching times) at
 1135 random from a uniform distribution centred on the posterior mean for that node in the
 1136 Beast tree and extending twice the length of the highest posterior density (HPD)
 1137 interval. For example, for a node with a 95% HPD interval of 2-3 Ma in our trees, the
 1138 uniform distribution was set to between 1.5 and 3.5 Ma. The HPD interval will capture
 1139 uncertainty under the selected phylogenetic and substitution models for the loci we
 1140 used, but we conduct our sensitivity analysis over a broader range to accommodate the
 1141 potential that the selected models and gene trees are inadequate. For cases where using
 1142 this approach meant that the lower bound was lower than 0, we assigned a value of
 1143 0.00001 Ma to the lower bound. We fitted the nine best models to the 100 datasets,
 1144 using five initial starting parameters for each model (total 4,500 optimisations). We
 1145 found that parameter estimates across the 100 datasets do not differ strongly from
 1146 those in the main dataset (Table XX). Importantly, model selection was unaffected, with
 1147 the M19 model being selected for all 100 datasets. This is because a lot of the
 1148 information used for model selection is coming from the other sources of information
 1149 DAISIE uses (island age, number of species, endemism status) rather than
 1150 colonisation/branching times.

1151

1152 **Table XX** – Parameter estimates of the M19 model obtained by fitting DAISIE to 100 datasets with
 1153 varying colonisation times and branching times chosen randomly chosen from an extended HPD
 1154 distribution. Compared to the estimates obtained in the main dataset. The M19 was selected as the
 1155 preferred model for all 100 datasets.

Parameter	Main dataset	100 datasets with variable ages	
		2.5 percentile	97.5 percentile
λ^c_0	0.040	0.035	0.043
d_0	0.026	0.025	0.027
μ_0	1.946	1.840	2.290
x	0.150	0.150	0.172
γ_0	67.22	60.99	69.99
α	0.294	0.291	0.297
λ^a_0	0.059	0.028	0.075
β	0.383	0.355	0.495

1156

1157 To account for uncertainty in **rates of molecular evolution**, we repeated all
 1158 BEAST dating analyses for markers that were not *cyt-b* using 1) the Weir and Schluter⁵⁰
 1159 *cyt-b* rate (dataset D1) and 2) marker specific rates estimated by Lerner et al.⁴¹, which

1160 are also widely used in the literature (dataset D2). Although the trees dated using the
1161 Lerner et al. rates provide younger ages, we found that the DAISIE results were very
1162 similar using either approach (same model preferred and similar parameters).
1163 Therefore, in the main text we only discuss the results of analyses of D1, i.e. applying the
1164 *cyt-b* rate of Weir and Schluter to all markers.

1165 For some taxa we did not use the stem age as the estimate of colonisation time,
1166 and instead used alternative nodes (see ‘Colonisation and branching times’ section). To
1167 test whether our **choice of nodes** affects our main conclusions, we recoded all such
1168 taxa by extracting the stem ages, and used these ages as an upper bound for
1169 colonisation (DAISIE MaxAge option). We fitted all 28 models to this new dataset (D3)
1170 and found that the M19 model is preferred and that the parameters vary only slightly
1171 from those of the main analysis. We therefore conclude our results are robust to the
1172 node selection approach.

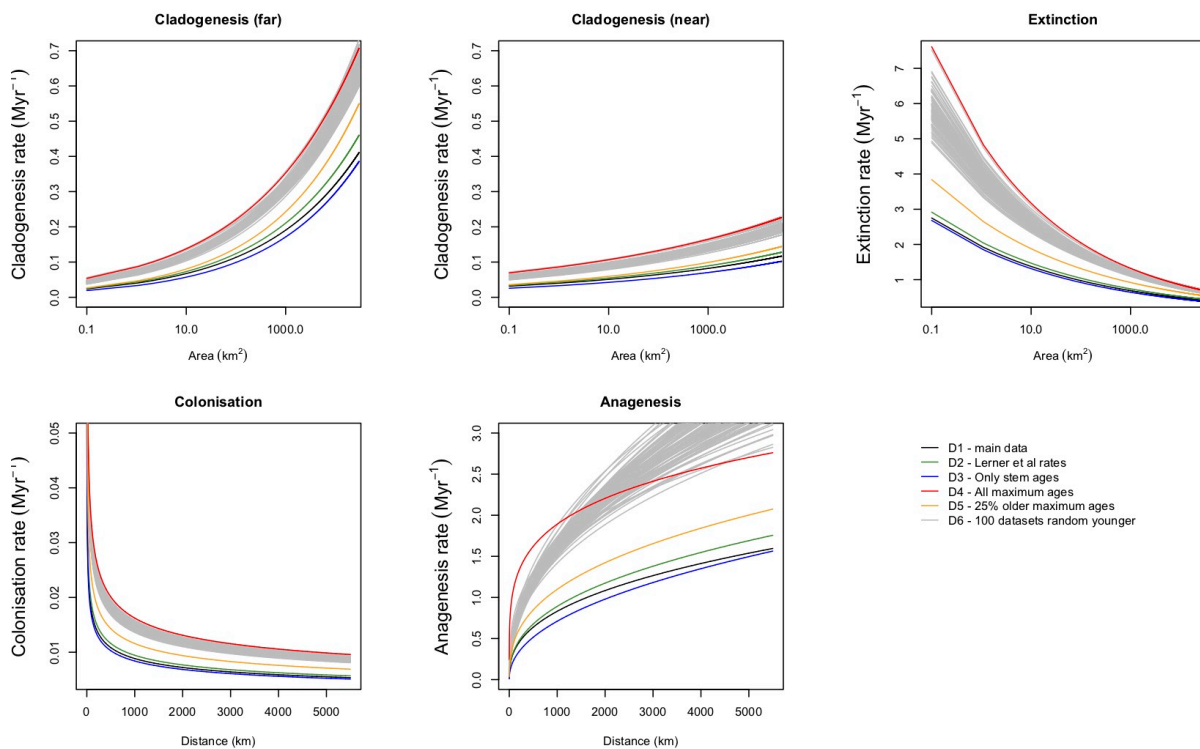
1173 If **extinction has been high on the mainland**, or if we failed to **sample the**
1174 **closest relatives** of the island taxa, this could lead to an overestimation of colonisation
1175 times when using the stem age as the precise time of colonisation. To investigate how
1176 this could have influenced our results, we ran analyses of datasets where we allow
1177 colonisation to have happened at any time since the stem age (i.e. the time of divergence
1178 from the nearest relative of the taxon on the mainland). For this we use the DAISIE
1179 options Endemic_MaxAge or NonEndemic_MaxAge, which integrate over all possible
1180 ages between the given maximum age and the present (or the first branching event
1181 within the archipelago for cases where cladogenesis has occurred). We repeated this
1182 analysis coding all stem ages as maximum ages (D4), or coding only the 25% older stem
1183 ages as maximum ages (to account for the fact that on older stems there is potential for
1184 there to be more bias) (D5). We also ran analyses on 100 datasets (D6) for which we
1185 assigned younger ages by randomly selecting a value between the stem age and the
1186 present (or crown age for cladogenetic groups). For all these datasets (D4-D6) we found
1187 that the same model (M19) is preferred, but the initial values of the biogeographical
1188 rates (cladogenesis, extinction, colonisation and anagenesis) are estimated to be higher
1189 than in the main dataset. Importantly, the exponent hyperparameters are similar to
1190 those in the main dataset. Thus, our main findings are robust to the colonisation times
1191 potentially being younger than those in our main dataset.

1192 The ML parameters of the M19 model for datasets D1 to D6 are given in table XX
 1193 and the resulting area and isolation dependencies are plotted in figure XX.

1194 **Table XX** – Parameter estimates of the M19 model obtained by fitting DAISIE to datasets with different
 1195 colonisation times and branching times.

Dataset	Characteristics	Cladogenesis			Extinction		Colonisation		Anagenesis	
		λ^0	γ	$d0$	μ_0	x	γ_0	α	λ^0	β
D1	Main dataset (cyt- <i>b</i> molecular rate applied to all markers)	0.040	0	0.027	1.946	0.150	67.256	0.294	0.059	0.383
D2	Molecular rates from Lerner et al applied to non-cyt- <i>b</i> markers	0.043	0	0.027	2.070	0.149	73.460	0.297	0.057	0.398
D3	Same as D1 but choosing stem ages for all colonisation times	0.033	0	0.028	1.877	0.155	63.106	0.292	0.029	0.463
D4	Same as D1 but using all ages as a maximum age	0.086	0	0.024	4.914	0.190	137.196	0.309	0.408	0.222
D5	Same as D1 but considering the 25% oldest stem ages in the dataset as maximum ages	0.046	0	0.028	2.687	0.155	94.449	0.304	0.083	0.373
D6	Same as D1 but with all ages younger (100 datasets, 2.5 and 97.5 percentile)	0.06 - 0.08	0	0.02 - 0.03	3.47 - 4.48	0.16 - 0.19	113.15 - 132	0.30 - 0.31	0.04 - 0.12	0.38 - 0.52

1196



1197

1198 **Fig XX - Estimated relationships between island area and isolation and local island biogeography**
 1199 **parameters for five datasets differing in colonisation and branching times.** Dataset names explained
 1200 in Table XX. Isolation measured as distance to the nearest mainland (D_m). Under the M19 model,

1201 cladogenesis rate increases with both area and distance, and thus plots for more (far, 5,000 km) and less
1202 (near, 50 km) isolated islands are shown.

1203

1204

1205 **Sensitivity to archipelago selection and isolation metrics**

1206

1207 The results of the following sensitivity analyses are presented in Supplementary Data 3
1208 and the DAISIE R objects including these alternative datasets have been deposited in
1209 Mendeley.

1210 To test whether the **inclusion of both true archipelagos and single islands** in
1211 our dataset could affect the results, we repeated analyses excluding single island units
1212 and found that the same model is preferred. The estimated initial rate of cladogenesis
1213 (λ^c_0) is higher if we exclude single islands, but this parameter is not different from a
1214 distribution of parameters estimated from datasets generated using a stratified-random
1215 sampling of both archipelagos and single islands.

1216 **Alternative isolation metrics** to D_m have been shown to explain varying and
1217 often higher amounts of variation in species richness on islands⁶⁰. We tested two
1218 alternative metrics: distance to the nearest larger or equivalent sized-landmass (D_b),
1219 and the mean between D_m and D_b (metrics given in Supplementary Data 2). We found
1220 that the same DAISIE model with very similar parameters was preferred in both cases,
1221 and thus we used only the D_m metric, as this is more similar to the original model of
1222 MacArthur & Wilson.

1223 The **Mascarenes** (Mauritius Isl., Reunion and Rodrigues) are often treated as a
1224 single biogeographical unit in analyses. We chose to analyse them as independent units
1225 because a) the distance between islands is much greater than our threshold for
1226 archipelago definition (more than 500 km between Mauritius Isl. and Rodrigues; more
1227 than 170 km between Reunion and Mauritius Isl.); b) only two species of our target
1228 group are shared between the islands (*Terpsiphone bourbonnensis* found in Mauritius
1229 Isl. and Reunion; and *Psittacula eques* found in Mauritius Isl. and extirpated from
1230 Reunion), suggesting low connectivity; c) while there are three clades whose branching
1231 events took place within the Mascarenes (*Coracina*, *Pezophaps/Raphus* and *Zosterops*),
1232 the remaining species result from independent colonisations suggesting that the three
1233 islands behave mostly as three different biogeographical units. We nevertheless ran an
1234 analysis treating the islands as a single archipelagic unit and found that the same model

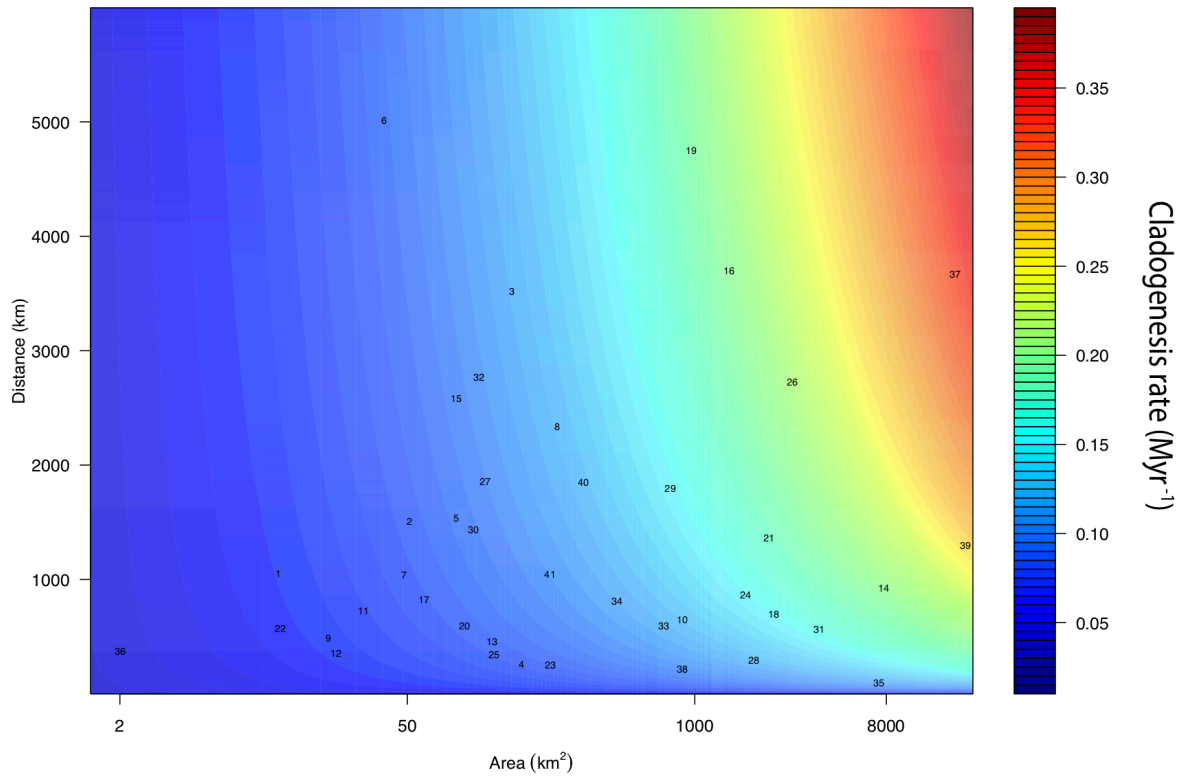
1235 was preferred and with similar parameter estimates, and we therefore discuss only the
1236 results treating them as separate.

1237 **Extended Data Figures and Tables**

1238

1239 Extended Data Figures 1 – 5

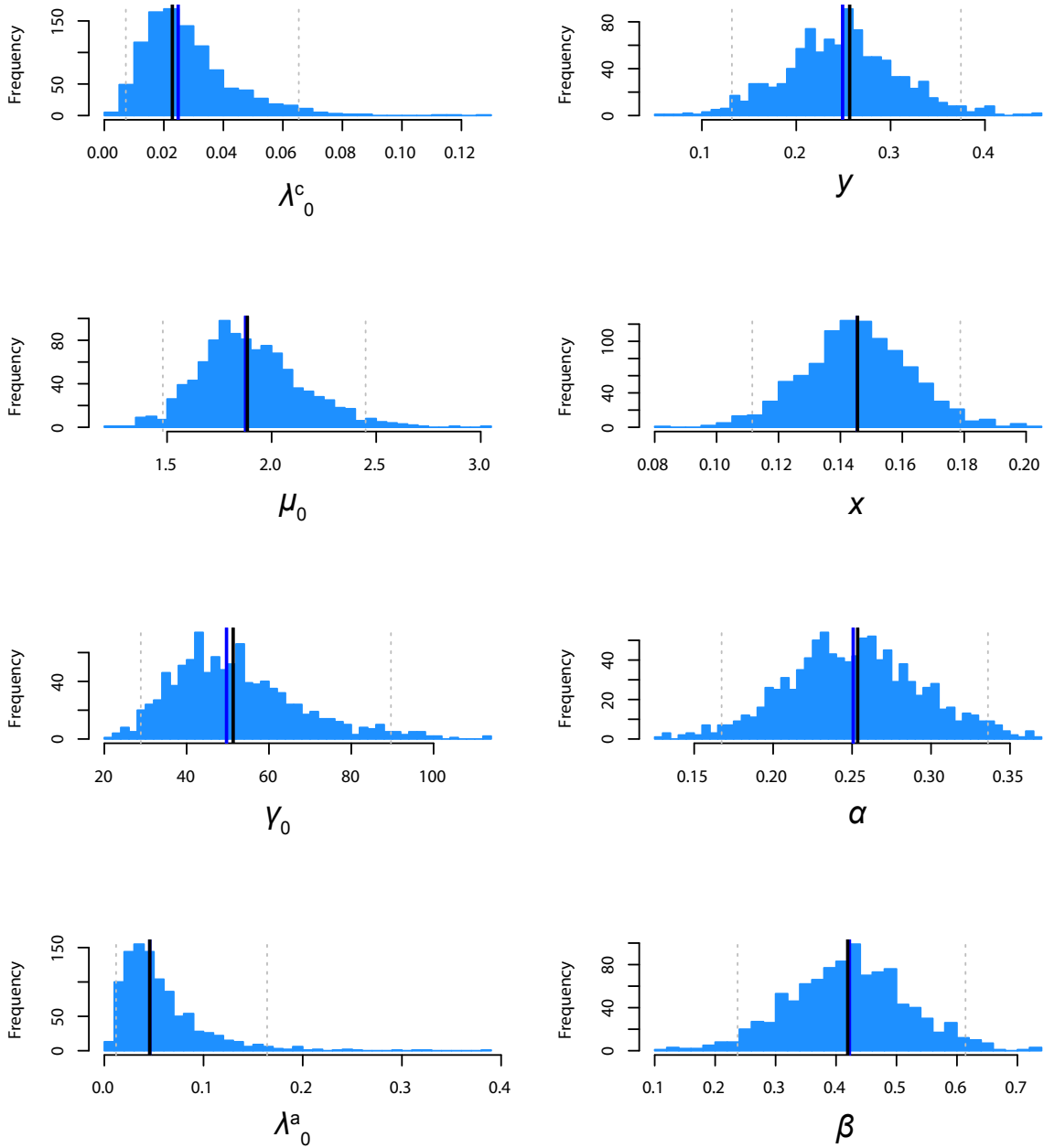
1240 Extended Data Tables 1 – 5



1241

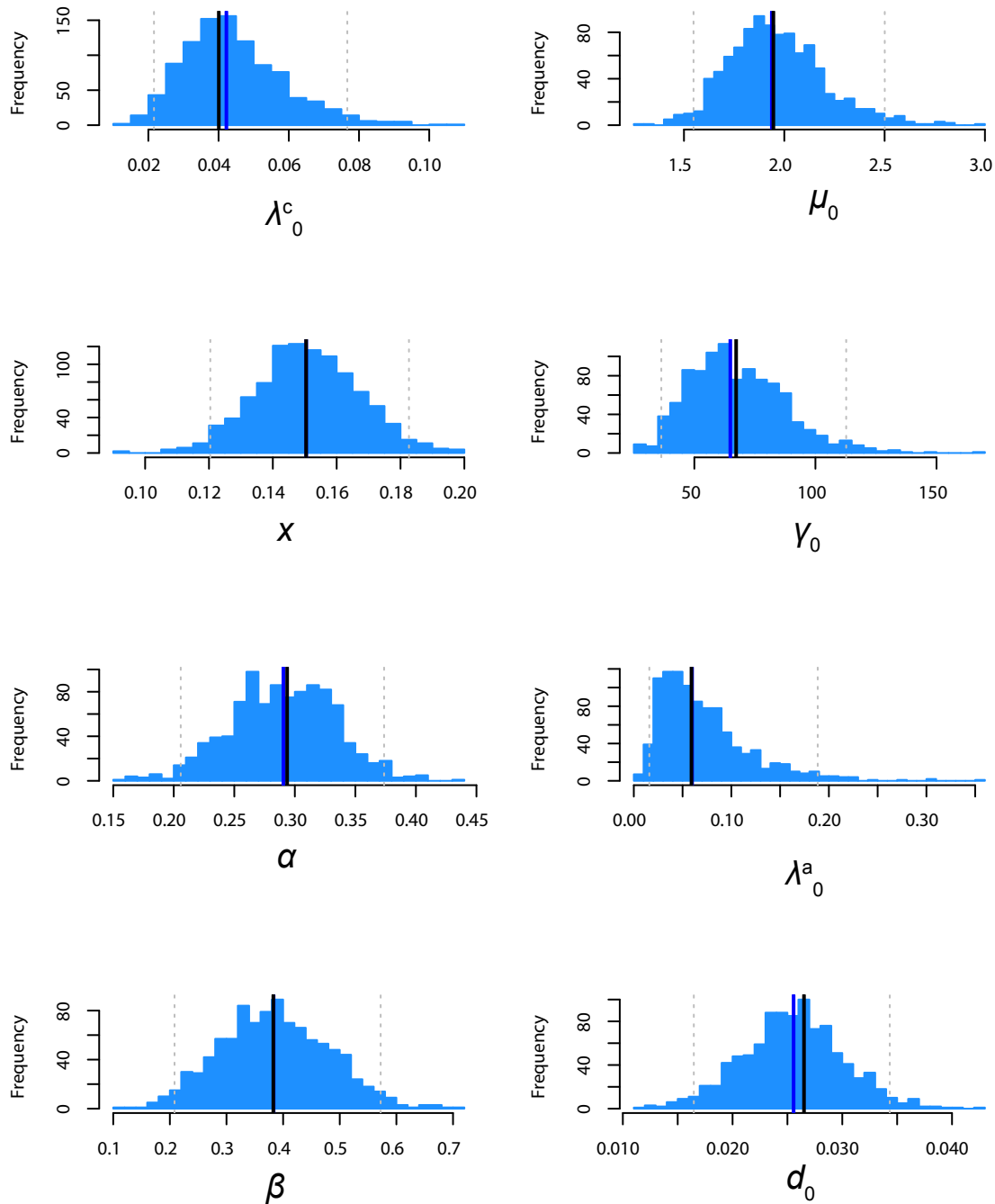
1242 **Extended Data Figure 1 | Variation of cladogenesis with isolation and area.**
 1243 Contour plot showing how the local rate of cladogenesis varies with area and distance
 1244 from the nearest mainland (D_m) assuming the maximum likelihood global
 1245 hyperparameters of the M19 model (equations describing the relationships given in
 1246 Supplementary Information Table 1). Numbers correspond to the archipelago numbers
 1247 from Fig. 1, and show the local cladogenesis rates for each of the archipelagos in our
 1248 dataset.

1249



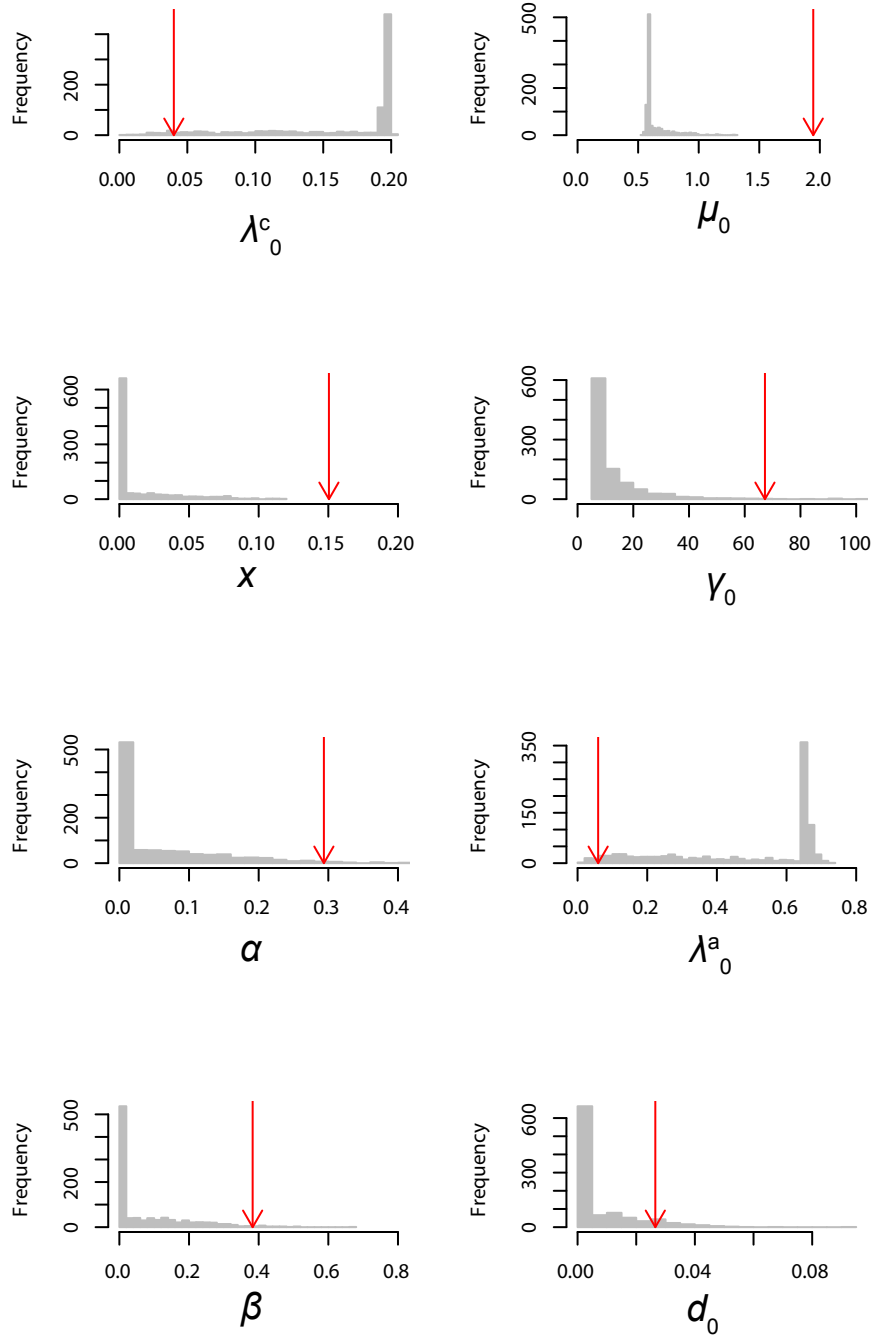
1250
 1251
 1252
 1253
 1254
 1255
 1256
 1257

Extended Data Figure 2 | Bootstrap precision estimates of the parameters of the M14 model. Parametric bootstrap analysis fitting the M14 model to 1,000 global data sets simulated with ML parameters of the M14 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 Information Table 1.



1258
 1259
 1260
 1261
 1262
 1263
 1264

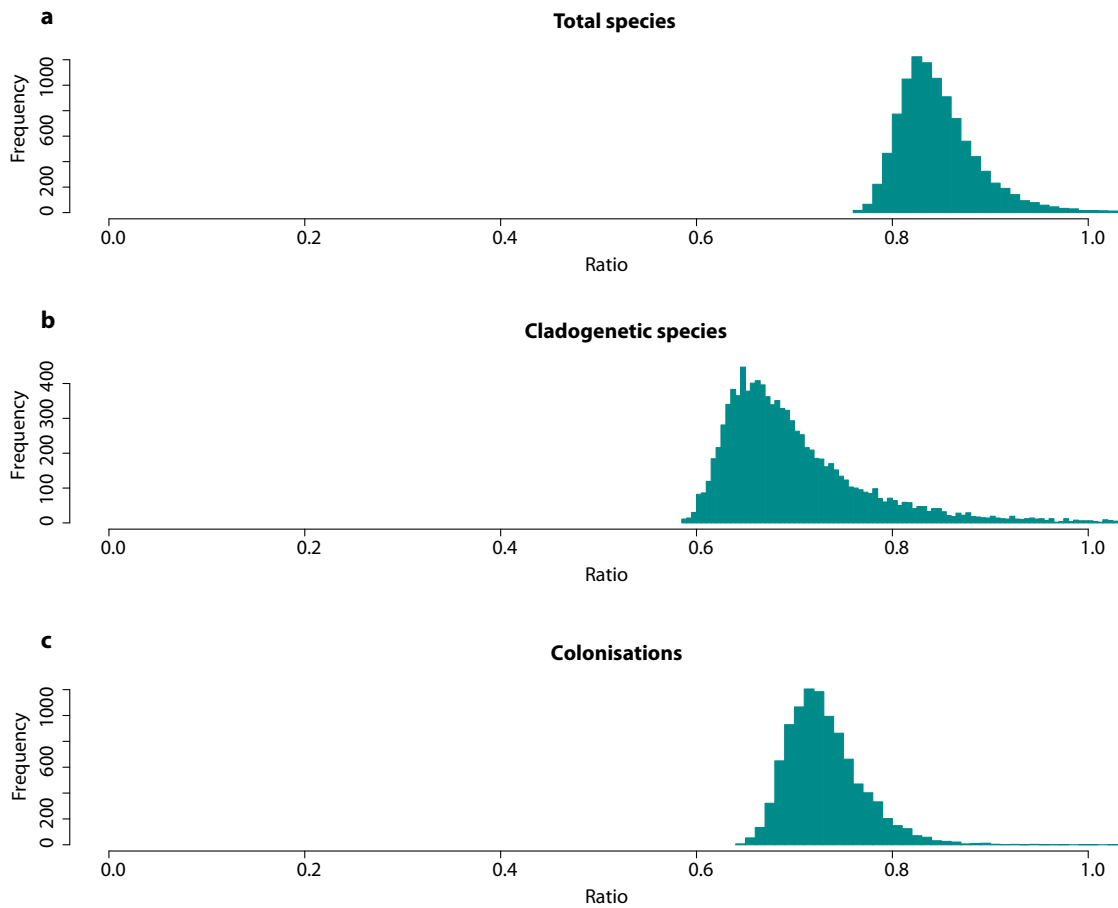
Extended Data Figure 3 | Bootstrap precision estimates of the parameters of the M19 model. Parametric bootstrap analysis fitting the M19 model to 1,000 global data sets simulated with ML parameters of the M19 model. Plots are frequency histograms of estimated parameters. Black lines show the median estimated values across all simulations and the blue lines the simulated values. Dashed lines show 2.5 – 97.5 percentiles. Parameters explained in Supplementary Information Table 1.



1265
1266

1267
1268
1269
1270
1271
1272
1273
1274
1275

Extended Data Figure 4 | 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 maximum likelihood 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 Information Table 1.



1276
1277
1278
1279
1280

Extended Data Figure 5 | Ratio of pseudo-R²-observed over pseudo-R²-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 maximum likelihood parameters.

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.

Archipelago	Area (km ²)	Distance nearest mainland (km)	Age (Ma)	Reference age	Number of species in focal group	Species sampled in phylogenies (percentage)*	Colonisations sampled in phylogenies (percentage)*
Aldabra Group	180	261	1	*	12	11 (91.67)	11 (91.67)
Ascension	91	1537	1	⁶¹	0	NA (NA)	NA (NA)
Azores	2387	1365	6.3	⁶²	17	13 (76.47)	13 (76.47)
Bermuda	53	1040	2	⁶³ †	5	2 (40)	2 (40)
Canary Islands	7493	96	21	⁶⁴	50	44 (88)	41 (87.23)
Cape Verde	4033	570	15.8	⁶⁵	10	10 (100)	10 (100)
Chagos	56.13	1510	0.0065	⁶⁶	0	NA (NA)	NA (NA)
Chatham	966	650	3	⁶⁷	14	13 (92.86)	13 (92.86)
Christmas Island	135	345	10	⁶⁸	4	2 (50)	2 (50)
Cocos (Costa Rica)	24	491	2.4	⁶⁹	4	2 (50)	2 (50)
Cocos (Keeling)	14.2	1054	0.003	⁷⁰	0	NA (NA)	NA (NA)
Comoros	2033	297	15	⁷¹	41	40 (97.56)	28 (96.55)
Fernando de Noronha	26	360	3.3	⁷²	3	1 (33.33)	1 (33.33)
Galápagos	7880	928	4	⁷³	27	26 (96.3)	8 (100)
Gough	91	2582	4	⁷⁴	1	1 (100)	1 (100)
Guadalupe	244	254	7	⁷⁵	11	10 (90.91)	10 (90.91)
Marianas (with Guam)	852	1800	15	‡	19	31 (60.78)	6 (75)
Hawaii	16624.6	3670	29.8	³⁶	51	6 (100)	5 (100)
Juan Fernández	99.67	600	5.8	⁷⁶	6	9 (81.82)	9 (81.82)
Lord Howe	14.55	571	6.9	⁷⁷	11	17 (89.47)	17 (89.47)
Madeira	798	600	18.8	⁷⁸	19	11 (57.89)	11 (57.89)
Marquesas	1063	4750	5.5	⁷⁹	19	12 (63.16)	7 (50)
Mauritius Isl.	1865	867	8.9	⁸⁰	15	9 (60)	9 (60)
New Caledonia	18576	1300	37	⁸¹	46	38 (82.61)	37 (82.22)
Niue	261.46	2340	2	⁸²	5	3 (60)	3 (60)
Norfolk	34.6	730	3.05	⁸³	14	11 (78.57)	11 (78.57)
Ogasawara	65	827	5	⁸⁴	10	5 (50)	5 (50)
Palau	488	815	20.1	⁸⁵	16	10 (62.5)	10 (62.5)
Pitcairn	42.8	5015	1.1	⁸⁶	8	3 (37.5)	2 (28.57)
Rapa Nui	163.6	3519	0.78	⁸⁷	2	0 (0)	0 (0)
Reunion	2512	700	5	⁸⁸	13	9 (69.23)	9 (69.23)
Rodrigues	109	1440	15	⁸⁹	9	4 (44.44)	4 (44.44)
Saint Helena	123.28	1856	14.5	⁹⁰	3	0 (0)	0 (0)
Samoa	3041	2730	13.5	⁹¹	22	16 (72.73)	16 (72.73)
São Tomé and Príncipe	964	219	30	⁹²	44	44 (100)	37 (100)
Selvagens	2.73	373	29	⁹³	1	1 (100)	1 (100)
Seychelles Inner	242.68	1048	64	³³	12	11 (91.67)	11 (91.67)
Society Islands	1577.8	3700	4.3	⁹⁴	18	7 (38.89)	7 (38.89)
Socorro	132	457	3.5	⁹⁵	7	7 (100)	7 (100)
Tonga	344.4	1850	41	⁹⁶	23	10 (43.48)	10 (43.48)
Tristan da Cunha	115.4	2770	18	⁹⁷	4	4 (100)	2 (100)

* ³⁵ 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.

Primer name	Sequence	Reference
New primers designed		
L-cytB_bird_0	TCAACRACTCCCTAATYGACCT	This paper
H-cytB_bird_2	AGRAYTACTCCTGTGTTTCARGTYTC	This paper
L-cytB_bird_2	GARACYTGAAACACAGGAGTARTYCT	This paper
H-cytB_bird_3	TAGKGGGTTGTTTGAGCCTGWTTTCGTG	This paper
L-cytB_bird_3	CACGAAWCAGGCTCAAACAACCC	This paper
H-cytB_bird_4	GGAGTAGTADGGGTGAAATGGRATTTT	This paper
H-cytB_bird_5	GGGTGTTCTACTGGTTGGCTKCC	This paper
L-cytB_bird_5	CCMCTCTCACAAAYCCTATTCTGA	This paper
L-cytB_human	TGAAACTTCGGATCCCTACTA	This paper
Published primers		
L14841	AAAAAGCTTCCATCCAACATCTCAGCATGATGAAA	98
L14995	GCCCCATCCAACATCTCAGCATGATGAAACTTCCG	99
L15308	GGC TAT GTC CTC CCA TGA GGC CAA AT	100
H15767	ATGAAGGGATGTTCTACTGGTTG	100
H15917	TAGTTGGCCAATGATGATGAATGGGTGTTCTACTGGTT	99
H16065	GAGTCTTCAGTCTCTGGTTTACAAGAC	101
L-cytB_Passer	CACAGGCCTAATTAAGCCTACCT	30
H-cytB_Passer	TTGARAATGCCAGCTTTGGGAG	30
L-cytB-Mot	CCAAATYGTTACAGGMCTCCTG	30
H-cytB-Mot	GGTGAATGAGGCTAGTTGCCCA	30

Extended Data Table 3 | Information on 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 Information Table 4.

Taxonomic group	Molecular marker(s)	Main source of sequences	Taxonomic group	Molecular marker(s)	Main source of sequences
<i>Acrocephalus</i>	cyt- <i>b</i>	-	<i>Moho</i>	cyt- <i>b</i>	-
Alaudidae (family)	cyt- <i>b</i>	-	<i>Monarcha</i>	cyt- <i>b</i>	-
<i>Alopecoenas</i> / <i>Gallinolumba</i>	ND2	-	<i>Motacilla</i>	cyt- <i>b</i>	-
<i>Anairetes</i>	cyt- <i>b</i>	-	<i>Myadestes</i>	cyt- <i>b</i>	-
<i>Anthus</i>	cyt- <i>b</i>	-	<i>Myiagra</i>	cyt- <i>b</i> + ND2	¹⁰⁶
<i>Aphrastura</i>	COI	-	<i>Myiarchus</i>	cyt- <i>b</i> + ND2	¹⁰⁷
<i>Bucanetes</i>	cyt- <i>b</i>	-	<i>Nigrita</i>	cyt- <i>b</i>	-
Buntings <i>Nesospiza</i> / <i>Rowettia</i>	cyt- <i>b</i>	-	<i>Onychognathus</i>	ND2	-
<i>Carduelis</i>	cyt- <i>b</i>	-	<i>Passer</i> / <i>Petronia</i>	cyt- <i>b</i>	-
<i>Chasiempis</i> / <i>Elepaio</i>	ND2	-	<i>Petroica</i>	cyt- <i>b</i>	-
<i>Chaunoproctus</i>	ND2	-	<i>Phylloscopus</i>	cyt- <i>b</i>	-
<i>Cinnyris notata</i>	ATP6	¹⁰²	<i>Pipilo</i>	cyt- <i>b</i>	-
<i>Cisticola</i>	cyt- <i>b</i>	-	<i>Pomarea</i>	cyt- <i>b</i>	-
<i>Clytorhynchus</i>	cyt- <i>b</i>	-	<i>Prinia</i>	cyt- <i>b</i>	-
<i>Coccyzus</i>	cyt- <i>b</i>	-	<i>Progne</i>	cyt- <i>b</i>	-
<i>Colaptes</i>	cyt- <i>b</i>	-	Psittaciformes	cyt- <i>b</i>	-
Columbiformes (order)	cyt- <i>b</i>	-	<i>Pyrrhocorax</i>	cyt- <i>b</i>	-
<i>Copsychus</i>	cyt- <i>b</i>	-	<i>Pyrrhula</i>	cyt- <i>b</i>	-
<i>Corvus</i>	cyt- <i>b</i>	-	<i>Regulus</i>	cyt- <i>b</i>	-
<i>Crithagra</i> / <i>Serinus</i>	cyt- <i>b</i>	-	<i>Saxicola</i>	cyt- <i>b</i>	-
Cuculiformes (order)	cyt- <i>b</i>	-	<i>Sephanoides</i>	cyt- <i>b</i>	-
<i>Cyanistes</i>	cyt- <i>b</i>	-	<i>Setophaga</i>	cyt- <i>b</i>	-
<i>Cyanolanius</i>	cyt- <i>b</i>	-	<i>Setophaga petechia</i>	ND2 + ATP8 + CR	¹⁰⁸
<i>Dendrocopos</i>	cyt- <i>b</i>	-	<i>Sitta</i>	cyt- <i>b</i>	-
<i>Dicrurus</i>	cyt- <i>b</i>	-	<i>Strepera</i>	cyt- <i>b</i>	-
<i>Dumetella</i>	ND2	-	<i>Sturnus</i>	ND2	-
<i>Emberiza</i>	cyt- <i>b</i>	-	Sunbirds (family)	cyt- <i>b</i>	-
<i>Erithacus</i>	cyt- <i>b</i>	-	<i>Sylvia</i>	cyt- <i>b</i>	-
<i>Estrilda</i> / <i>Erythrura</i>	cyt- <i>b</i>	-	<i>Terpsiphone</i>	cyt- <i>b</i>	-
Finches, Galápagos Cocos	cyt- <i>b</i> + Multiple	¹⁰³	<i>Troglodytes</i> / <i>Thryomanes</i>	ND2	¹⁰⁹
<i>Foudia</i>	ATP8 + ATP6 + ND3	¹⁰⁴	<i>Turdus</i>	cyt- <i>b</i>	-
<i>Fregilupus</i>	ND2	-	<i>Upupa</i>	cyt- <i>b</i>	-
<i>Fringilla</i>	cyt- <i>b</i>	-	<i>Vidua</i>	cyt- <i>b</i>	-
<i>Haemorhous</i>	cyt- <i>b</i>	-	Weavers (family)	cyt- <i>b</i>	-
Hawaiian Honeycreepers	cyt- <i>b</i>	⁴¹	<i>Zosterops</i> (Indian, Atlantic)	cyt- <i>b</i> , ND3	¹¹⁰
<i>Horornis</i>	cyt- <i>b</i>	-			
<i>Humblotia</i>	cyt- <i>b</i>	-			
<i>Hypsipetes</i>	ND3	¹⁰⁵			
<i>Lamprotornis</i>	ND2	-			
<i>Lanius</i>	cyt- <i>b</i>	-			
<i>Leptosomus</i>	cyt- <i>b</i>	-			
<i>Lonchura</i>	cyt- <i>b</i>	-			
<i>Loxia</i>	cyt- <i>b</i>	-			
<i>Microeca</i> / <i>Eopsaltria</i>	cyt- <i>b</i>	-			
<i>Mimus</i>	ND2	-			

1 **Extended Data Table 4** | Information on the previously published dated trees used.

2

Taxonomic group	Source	Molecular markers	Calibration method
<i>Calypte</i>	¹¹¹	Multiple	Molecular rate
<i>Cinclodes</i>	¹¹²	3 mtDNA and 3 nuclear	Biogeographical
Corvides	¹¹³	Multiple	Fossils
<i>Corvus moriorum</i>	¹¹⁴	Mitogenome	Fossils
<i>Ducula</i>	¹¹⁵	ND2, COI, ND3, nuclear	Secondary & Biogeography
<i>Junco</i>	¹¹⁶	ND2 + CR + COI + ATP + nuclear	Molecular rate
Meliphagides (infraorder)	¹¹⁷	mtDNA and nuclear	Fossils & Secondary
<i>Nesillas</i>	¹¹⁸	ND2	Molecular rate
<i>Ptilinopus</i>	¹¹⁹	ND2, COI, ND3, nuclear	Secondary & Biogeography
<i>Pyrocephalus</i>	¹²⁰	cyt- <i>b</i> , ND2, nuclear	Molecular rate
<i>Zosterops</i> (Pacific)	¹²¹	cyt- <i>b</i> , ND2, ND3, ATPase	Molecular rate

3

4 **Extended Data Table 5** | Maximum likelihood estimates and 95% confidence intervals
 5 of the parameters of the two best models. Confidence intervals obtained from the
 6 bootstrap analyses. Parameter symbols explained in Supplementary Information Table
 7 1.

8

Model	Cladogenesis		Extinction		Colonisation		Anagenesis	
	λ_0^c	γ	μ_0	x	γ_0	α	λ_0^a	β
M14	0.023 (0.01 - 0.07)	0.26 (0.13 - 0.37)	1.88 (1.48 - 2.45)	0.15 (0.11 - 0.18)	51.30 (28.86 - 89.64)	0.25 (0.17 - 0.34)	0.05 (0.01 - 0.16)	0.42 (0.24 - 0.61)
M19	0.04 (0.022 - 0.077)	0.027 (0.016 - 0.034)	1.95 (1.55 - 2.50)	0.15 (0.12 - 0.18)	67.26 (36.35 - 112.71)	0.29 (0.21 - 0.37)	0.059 (0.02 - 0.19)	0.38 (0.21 - 0.57)

9