1	Title	page
---	-------	------

2	Article title: Taxon Cycle predictions supported by model-based inference in Indo-Pacific trap-
3	jaw ants (Hymenoptera: Formicidae: Odontomachus).
4	Authors: Pável Matos-Maraví ^{a,b,c,d} , Nicholas J. Matzke ^{e,f} , Fredrick J. Larabee ^{g,h} , Ronald M.
5	Clouse ⁱ , Ward C. Wheeler ⁱ , D. Magdalena Sorger ^j , Andrew V. Suarez ^h , Milan Janda ^{a,k}
6	Institutions:
7	^a Institute of Entomology, Biology Centre CAS, Branisovska 31, CZ-37005 Ceske Budejovice,
8	Czech Republic
9	^b Department of Zoology, Faculty of Science, University of South Bohemia, Ceske Budejovice,
10	Czech Republic
11	^c Department of Biological and Environmental Sciences, University of Gothenburg, Göteborg,
12	Sweden
13	^d Gothenburg Global Biodiversity Centre, Göteborg, Sweden
14	^e Division of Ecology and Evolution, Research School of Biology, The Australian National
15	University, ACT, Australia
16	^f School of Biological Sciences, The University of Auckland, New Zealand
17	^g Department of Entomology, National Museum of Natural History, Smithsonian Institution,

18 Washington, DC, USA

19	ⁿ Department of Entomology and Department of Animal Biology, University of Illinois, Urbana-
20	Champaign, Urbana, IL, USA

- ⁱDivision of Invertebrate Zoology, American Museum of Natural History, 200 Central Park West,
 New York City, NY, USA
- ^jDepartment of Applied Ecology, North Carolina State University, Raleigh, NC, USA; W.M.
- 24 Keck Center for Behavioral Biology, North Carolina State University, Raleigh, NC 27695, USA;
- 25 Research & Collections, North Carolina Museum of Natural Sciences, Raleigh, NC 27601, USA
- 26 ^kLaboratorio Nacional de Análisis y Síntesis Ecológica, ENES, UNAM, Morelia, Mexico
- 27 Keywords: biogeography; diversification; Formicidae; insect; Melanesia; taxon cycle
- 28 Corresponding author: Pável Matos-Maraví; Institute of Entomology, Biology Centre CAS,
- 29 Branisovska 31, CZ-37005 Ceske Budejovice, Czech Republic; Fax: +420385310354; Electronic
- 30 mail: pavelm14@gmail.com
- 31 **Running title:** The Taxon Cycle revisited

32 Abstract

Non-equilibrium dynamics and non-neutral processes, such as trait-dependent dispersal, are often 33 missing from quantitative island biogeography models despite their potential explanatory value. 34 35 One of the most influential non-equilibrium models is the taxon cycle, but it has been difficult to test its validity as a general biogeographical framework. Here, we test predictions of the taxon-36 cycle model using six expected phylogenetic patterns and a time-calibrated phylogeny of Indo-37 Pacific Odontomachus (Hymenoptera: Formicidae: Ponerinae), one of the ant genera that E.O. 38 Wilson used when first proposing the hypothesis. We used model-based inference and a newly 39 40 developed trait-dependent dispersal model to jointly estimate ancestral biogeography, ecology (habitat preferences for forest interiors, versus "marginal" habitats, such as savannahs, shorelines, 41 disturbed areas, etc), and the linkage between ecology and dispersal rates. We found strong 42 evidence that habitat shifts from forest interior to open and disturbed habitats increased 43 macroevolutionary dispersal rate. In addition, lineages occupying open and disturbed habitats can 44 45 give rise to both island-endemics re-occupying only forest interiors and taxa that re-expand geographical ranges. The phylogenetic predictions outlined in this study can be used in future 46 work to evaluate the relative weights of neutral (e.g., geographical distance and area) and non-47 48 neutral (e.g., trait-dependent dispersal) processes in historical biogeography and community ecology. 49

50 Introduction

Islands are natural laboratories for ecologists and evolutionary biologists. Their discrete nature 51 52 can facilitate the reconstruction of evolutionary and biogeographic histories of resident flora and 53 fauna, especially compared to continental systems. Despite major advances in island biogeography following the publication of its most influential model more than a half-century 54 55 ago (MacArthur & Wilson, 1967), there remain many open questions that require multidisciplinary efforts to address (Patiño et al., 2017). For example, the reconciliation of 56 ecology and evolutionary history can improve our understanding of the impact of ecological or 57 biological traits on island biogeographic patterns (Valente et al., 2015; Whittaker et al., 2017). 58 Non-neutral processes, such as trait-modulated dispersal and biotic innovations, often remain 59 neglected in island biogeography models. Perhaps, the most influential non-equilibrium model 60 that explicitly incorporates non-neutral processes into an evolutionary context is the taxon cycle 61 (Wilson, 1959a, 1961; Ricklefs & Cox, 1972; Ricklefs & Bermingham, 2002). The taxon cycle 62 proposes that dispersal and ecological adaptation interact with competition, range contraction, 63 64 and extinction to explain observed species distribution. The historical narrative of the taxon cycle 65 describes the diversity dynamics of island biota in a sequence of stages (Fig. 1). In Stage I, species expand their geographical ranges across archipelagos aided by ecological release and an 66 67 adaptive shift to occupy "marginal" habitats. Wilson's (1959a, 1961) definition of "marginal" habitat includes littoral zones, open habitats, and disturbed environments; for simplicity and to 68 avoid any misinterpretation, we use open habitat to refer to "marginal" habitat throughout the 69 paper. In Stage II, extinction and population differentiation leave patchy species distributions 70 71 across archipelagos. In Stage III, single-island endemics evolve the potential to either further

72	radiate within islands or re-initiate the taxon cycle by entering into Stage I. Empirical studies
73	typically seek patterns left by hypothetical taxon cycles on island communities, but these have
74	usually relied on qualitative interpretations of distributional data, including altitudinal and habitat
75	occurrences (Cook et al., 2008; Ricklefs & Bermingham, 2008; Dexter, 2010; Economo &
76	Sarnat, 2012; Jønsson et al., 2014; Economo et al., 2015). Revisiting the taxon cycle is timely
77	with the addition of new probabilistic models for historical and island biogeography that allow
78	traits or geographic range to influence dispersal or diversification rates (Goldberg et al., 2011;
79	Sukumaran et al., 2016; Sukumaran & Knowles, 2018; Klaus, Matzke et al., in review).
80	In this study, we test phylogenetic predictions of the taxon cycle using model-based inference
81	(Burnham & Anderson, 2002) with recently developed computational models. We apply these
82	models to a comprehensive time-calibrated phylogeny of the ant genus Odontomachus, one of the
83	taxa E.O. Wilson used in his original articulation of the taxon-cycle model (Wilson, 1959a,
84	1961). Odontomachus occurs in tropical and subtropical biomes worldwide, and is particularly
85	diverse in the insular landscape of the Indo-Pacific region where 31 out of 68 described species in
86	the genus occur. Based on extant distributions and taxonomic affinities, it was believed that most
87	Indo-Pacific ant lineages originated in tropical SE Asia and Australia (the "source regions" sensu
88	Wilson, 1959a). However, New Guinea may also be a primary source of Odontomachus (Wilson,
89	1959a, 1961) and other ants in the Indo-Pacific region (Lucky & Sarnat, 2010). Recent
90	biogeographical analyses suggest that continental SE Asian Odontomachus did not spread to New
91	Guinea (instead a single dispersal event from the New World has been suggested), and Australian
92	species form a derived monophyletic subgroup nested within New Guinean lineages (Larabee et
93	al., 2016). Expanding taxa in Stage I might further differentiate into Stage II and Stage III mostly

on the largest Indo-Pacific archipelagos such as Fiji, the Philippines or Borneo, due to the greater
amount of habitat and elevational gradients with increasing island area size (Wilson, 1959a,
1961; Economo & Sarnat, 2012).

97 Computational models can readily be applied to some aspects of the taxon cycle. Theoretically, the most desirable test of the taxon-cycle would use a complete computational model that 98 explicitly integrates all ecological and evolutionary processes proposed in Wilson's verbal model 99 of the taxon cycle, including speciation, extinction, dispersal, competition, biotic innovations, and 100 interactions among these processes. However, such a computational model does not vet exist. In 101 102 particular, accurate inference of rates of lineage extinction, using only phylogenies of living taxa, 103 is likely to be difficult even in simple models, let alone complex models integrating many 104 processes. However, available computational models can estimate the rates of processes that are most accessible with phylogenies of living taxa, transition rates for an ecological trait such as 105 habitat preference, and biogeographic dispersal rates. These models also provide estimates of 106 107 ancestral geographic ranges and ecology, as well as the approximate timing of when changes 108 occurred. While these inferences do not cover all aspects of the taxon-cycle model, they enable testing of key aspects of it. In addition, a key prediction of the taxon cycle involves the 109 110 relationship between ecological shifts and range expansion. Here, we combine an ecological trait 111 (habitat association) and biogeography into a single model using a trait-dependent dispersal model, where ancestral ecology and historical biogeography are jointly inferred on the phylogeny 112 113 of Indo-Pacific Odontomachus, and where trait state can influence dispersal rate (Klaus, Matzke et al., in review). In summary, using computational inference models of trait evolution, 114

geographic range evolution, or both simultaneously, we test six predictions stemming from thetaxon-cycle model (Table 1):

117 Prediction 1: Importance of range-expansion dispersal. Stage I range expansion is a populationlevel process, and Stage II and Stage III range contraction trigger speciation at a 118 macroevolutionary level. Evidence to support this prediction at Stage I should be anagenetic 119 120 range expansion (i.e., along branches in a phylogenetic tree), and at Stage II and III should be within-area cladogenesis (similar to "subset-sympatry" in Fig. 1). Given the phylogeny of 121 Odontomachus, model-based inference would support parametric biogeographical models that 122 123 include widespread ancestral taxa at cladogenesis and dominance of within-area speciation, as 124 opposed to vicariance and jump-dispersal speciation.

Prediction 2: Timing of dispersal events. Stage I initiates with expanding ranges from large and 125 diverse areas (i.e., source regions). The total number of dispersal events among Indo-Pacific 126 127 archipelagos should reveal that expansion out of New Guinea, the assumed primary source region for Indo-Pacific Odontomachus based on morphological affinities (Wilson, 1959b; Brown, 1976), 128 129 patterns of distribution (Wilson, 1959a, 1961), and previous biogeographical analyses (Larabee et 130 al., 2016), is older than range expansions back into New Guinea from the Pacific region. Notably, this prediction does not contradict young dispersal out of source regions; instead, we 131 132 acknowledge that dispersal events out of/into source regions continue to the present. The timing of dispersal events on the phylogeny of Odontomachus can be estimated using biogeographical 133 stochastic mapping; when this is done, the ages of all dispersal events out of New Guinea should 134 be significantly older than dispersal into New Guinea. 135

Prediction 3: Rates of transition in habitat preference. The transition rate from open habitat (e.g., 136 137 littoral and disturbed environments) to undisturbed forest interior should be higher than in the reverse direction, because taxa occupying open and disturbed habitats encounter two evolutionary 138 139 fates: 1) extinction or 2) habitat shift and survival in the forest interior. Evidence to support this 140 pattern includes most extant species occupying forest interior habitat, while the open habitat state would be estimated at deeper nodes in the phylogeny. A trait-dependent dispersal model that 141 jointly infers biogeography and habitat state transitions would estimate state transition rates from 142 open habitat to forest interior to be higher than from forest interior to open habitat. 143

144 Prediction 4: Dispersal rates in each habitat. Range expansion by Stage I taxa is linked to their ecological preference for open habitat, while taxa occupying only forest interior have negligible 145 146 dispersal across archipelagos. Higher macroevolutionary dispersal probabilities are expected for taxa associated with open habitat, and extant taxa occupying open habitat should be 147 geographically widespread (Stage I). Therefore, a trait-dependent dispersal model should accrue 148 149 higher likelihood than a model where dispersal rates are the same between open versus forest 150 interior habitats, and the open habitat state should increase dispersal probability among archipelagos. 151

152 Prediction 5: Timing of dispersal events. Stage I taxa disperse readily due to their preference for 153 open habitats, and are thus not dependent on external environmental drivers such as sea-level 154 drop and subsequent increased landmass size and connectivity. Therefore, dispersal events by 155 Stage I taxa should be asynchronous across clades, whereas if dispersal was caused by a major 156 external driver, approximately simultaneous dispersal events should be expected. Given the time157 calibrated phylogeny of *Odontomachus*, ancestral character state shifts to open habitat (Stage I)158 should be estimated at different times.

Prediction 6: Widespread taxa should have broad ecological niches. Stage I taxa with broad geographical ranges belong to lineages having broad ecological preferences, including associations with open and disturbed habitat. Widespread Stage I taxa would belong to clades dominated by single-area endemics having preference for forest interior. Ancestral state inference should indicate that clades that contain widespread taxa have a high probability for the open habitat state at the crown node.

165 Material and methods

166 *The study taxon*

167 The trap-jaw ant genus *Odontomachus* has recently been the subject of taxonomic reviews

168 (Yoshimura et al., 2007; Fisher & Smith, 2008; Sorger & Zettel, 2011; MacGown et al., 2014;

169 Satria et al., 2015) and higher-level molecular systematic studies (Larabee et al., 2016).

170 *Odontomachus* is a monophyletic genus comprised of 68 valid extant species (Schmidt &

171 Shattuck, 2014; Satria et al., 2015; AntWeb, 2017), which are unevenly distributed across

172 continents. The Oriental region (i.e., tropical and subtropical East Asia extending through the

173 Malay Archipelago region west to the Wallace's Line) and the Australasian region (i.e. the Malay

174 Archipelago region east of Wallace's Line and Australia) together harbor the largest number of

175 *Odontomachus* species, with 31 valid taxa. The New World has 27 described species, whereas the

176 Afrotropics-Malagasy and the Asian Palearctic regions have only 3 and 7 species, respectively.

The lower species diversity found in the last two regions is not a sampling artifact, given recent
research efforts there (e.g., Fisher & Smith, 2008; Liu et al., 2015; Janicki et al., 2016).

179 Taxon sampling and molecular dataset

180 In the recent phylogeny of Larabee et al. (2016), the species-rich Malay Archipelago region lacked comprehensive taxonomic sampling, thus hindering a thorough phylogenetic test of the 181 182 taxon-cycle model on these ants. Larabee et al. (2016) studied 43 Odontomachus specimens each 183 of them representing one species, and only 17 individuals were from the Indo-Pacific. To overcome this, we conducted extensive sampling in the Indo-Pacific during 2002–2015, with 184 emphasis on Melanesia (Fig. S1 and Appendix S1). In addition, we included in our analyses 185 specimens deposited at the Museum of Comparative Zoology (MCZ) at Harvard University, the 186 Smithsonian National Museum of Natural History (USNMENT), and the CSIRO Tropical 187 Ecosystems Research Centre, Darwin, Australia (TERC). We sorted specimens primarily based 188 189 on morphology, using published taxonomic keys (Wilson, 1959b; Brown, 1976; Sorger & Zettel, 2011) and comparing most ant individuals directly to type collections at MCZ and at the USNM. 190 191 To further corroborate our morphological identifications and find taxa with substantial genetic 192 variation that may represent cryptic species, molecular-based species determination using a 193 multi-locus dataset was carried out only for the Indo-Pacific clade (Figs S3–S5). The photographs 194 of the specimen vouchers and associated distribution data are available at the public database 195 Ants of New Guinea (http://www.newguineants.org/).

We followed standard lab protocols for genomic DNA isolation, amplification, and Sanger
sequencing (Schultz & Brady, 2008; Schmidt, 2013; Clouse et al., 2015). We also retrieved

publicly available DNA sequences in GenBank of *Odontomachus* and its sister genus *Anochetus*. 198 199 We generated a molecular dataset of about 3.8 kb aligned gene sequences, encompassing the 200 protein-coding mitochondrial gene COI, and the nuclear gene fragments CAD (including one 201 intron), EF-1 α F1, LWR (including one intron), and *wingless*, as well as a fragment of the 202 ribosomal gene 28S. The alignment of the two intronic regions was straightforward within Odontomachus and the outgroup taxon Anochetus. Sequencing of both DNA strands was carried 203 out by Macrogen (South Korea) and the edition and alignment of sequences were conducted in 204 Geneious R7. All DNA sequences are available in GenBank and in BOLD (Barcode of Life) 205 206 under the Ants of Papua New Guinea (ASPNA) project (Appendix S1).

207 Phylogenetic analyses

208 Single-gene datasets and a concatenated dataset consisting of 93 taxa having at least three sequenced gene fragments were analyzed in MrBayes v3.2.3 (Ronquist et al., 2012). The best 209 210 partitioning strategy for the concatenated dataset was suggested by PartitionFinder v1.1.1 211 (Lanfear et al., 2012), using 18 data blocks considering each gene's coding position, parameters 212 set to branchlengths=linked (better likelihood than the unlinked option), model_selection=bic, 213 and search=greedy. Phylogenetic analyses in MrBayes were carried out through CIPRES (Miller et al., 2010), and consisted of two independent runs each for 50 million generations, sampling 214 215 every 5,000 generations. The nucleotide substitution scheme "mixed + Γ " was set to each partition, which permits the exploration of best-fitting reversible models in the MCMC sampling 216 (Huelsenbeck et al., 2004). The first 25% of sampled parameters were discarded as burn-in. We 217 checked that the final average standard deviation split frequencies were below 0.01, PSRFs were 218 219 approaching unity, ESS values were higher than 200, and log probabilities reached stationary

distribution. We summarized the MCMC sampled trees using the 50% majority rule consensusapproach.

222 We also ran Maximum Parsimony (MP) and Maximum Likelihood (ML) analyses on the same 223 concatenated dataset, with ML partitioned under the best-fit strategy as above. We used RAXML v8.1.11 (Stamatakis, 2014) to conduct the ML analyses under the "rapid bootstrapping" algorithm 224 225 with 1,000 iterations. We set the GTRGAMMA model for each partition and we searched for the 226 best-scoring tree using the command "-f a". We computed an extended majority rule consensus tree from the 1,000 bootstraps trees using the command "-J" in RAxML. The program TNT v1.5 227 228 (Goloboff et al., 2008) was used to conduct the MP analysis using "new technology search" 229 algorithms (Goloboff, 1999; Nixon, 1999) with the following parameters: RSS, CSS, and XSS 230 sectorial searches, Ratchet with 10 iterations, Drift with 10 cycles, and Tree Fusing with 10 231 rounds. Node support was assessed using standard bootstrapping with 1,000 pseudo-replicates.

232 Divergence time estimation

233 Five extinct species, two within Anochetus and three within Odontomachus, were used as fossil-234 based calibrations following the guidelines of Parham et al. (2012). All five fossils are well-235 diagnosed, and their affinities within infra-generic groups were proposed in the light of well-236 preserved, apomorphic characters (Table 2). The compression fossil of O. paleomyagra was found in the Bílina Mine coal seam, Most Basin, Czechia (Wappler et al., 2014), with an age of 237 about 20 Ma in the Burdigalian Stage, Early Miocene (Kvaček et al., 2004; Knor et al., 2013). 238 The remaining four extinct species were found as inclusions in Dominican amber (Baroni Urbani, 239 240 1980; de Andrade, 1994), but their ages were not accurately correlated with stratigraphic levels.

However, a mid-Miocene origin of Dominican amber has been proposed (Iturralde-Vinent & 241 242 MacPhee, 1996; Iturralde-Vinent, 2001; Grimaldi & Engel, 2005). We used the concatenated 243 multi-locus dataset with one specimen per species and fossil data to infer time-calibrated species 244 divergences using the fossilized birth-death (FBD) model (Heath et al., 2014). We enforced 245 monophyly of extant and extinct taxa (the latter with missing nucleotide data in the alignment) for the taxonomic groups shown in Table 2, and we set the following parameters: "diversification 246 rate" (Exponential; mean = 1.0), "sampling proportion" (Beta; alpha = 2.0, beta = 2.0), and 247 "turnover rate" (Uniform; from 0 to 1), whereas the proportion of sampling extant species was set 248 249 to 0.5 based on the species checklist for *Odontomachus* (AntWeb, 2017). We applied a soft prior 250 lognormal distribution for the root age, i.e., Anochetus and Odontomachus divergence: upper and lower 2.5% quantiles from 20.9 to 67.4 Ma (Larabee et al., 2016), with mean age of 31 Ma based 251 on previous fossil-calibrated Ponerinae phylogeny (Schmidt, 2013). Alternatively, we calculated 252 253 divergence times using a calibration-density approach with minimum clade ages modeled as 254 exponential distributions reflecting the fossils dates (Fig. S6). 255 The analysis was run in BEAST v2.3.1 (Bouckaert et al., 2014) using the lognormal uncorrelated Bayesian relaxed-clock (Drummond et al., 2006). Substitution models as suggested by 256 257 jModelTest v2.1.7 (Darriba et al., 2012) were unlinked across partitions, and MCMC approximations were run four independent times for 200 million generations, sampling every 258 20,000 generations, and discarding the first 25% sampled parameters in each run. We 259 260 summarized sampled trees into a maximum clade credibility (MCC) tree, after verifying 261 convergence and mixing of chains (ESS > 200). DNA alignment and time-calibrated phylogenies

- were deposited at TreeBase (Study ID 20232) and Dryad (DOI:
- 263 https://doi.org/10.5061/dryad.5542pr8).
- 264 Probabilistic inference of historical biogeography

265 Ancestral geographical ranges within the genus *Odontomachus* were inferred using the R v.3.4.2 266 package BioGeoBEARS v0.2.1 (Matzke, 2013a, 2014; R Core Team, 2017), with updates to the code as posted at http://phylo.wikidot.com/biogeobears. Ranges were inferred on the MCC tree 267 268 from BEAST under the FBD model. We used the following biogeographical base models in our analyses (Matzke, 2013b): 1) Dispersal-Extinction-Cladogenesis (DEC) (Ree et al., 2005; Ree & 269 270 Smith, 2008), 2) DIVA-like (a likelihood implementation of the biogeographical processes 271 assumed in DIVA, Ronquist, 1997), and 3) BayArea-like (a likelihood implementation of the 272 biogeographical processes allowed in BayArea, Landis et al., 2013). The main differences among the three biogeographical base models are related to the assumed scenarios at cladogenesis: 1) 273 274 vicariance (allowed only by DEC and DIVA-like models in slightly different ways; Matzke, 275 2013b); 2) within-area speciation, wherein 2a) both daughter lineages inherit the same ancestral 276 geographical range (allowed in all three base models) or 2b) one of the two daughter lineages 277 inherits only a subset of a wider distributional range of the parent node (allowed only in DEC; see Fig. 1). Only the BayArea-like model allows "range copying" (where a species occupying 278 279 multiple areas gives rise to two daughters copying the same widespread range, i.e., widespread sympatry). In terms of expectations (Prediction 1; Table 1), we do not foresee vicariance playing 280 a major role in the Odontomachus dataset, but we do for within-area speciation, either within the 281 same areas (e.g., many speciation events only within the New World or New Guinea), or within a 282 283 subset of the parental geographical range (e.g., rising of island endemics from ancestral

widespread taxa in the Indo-Pacific). All of these models effectively assume that a Yule-process
generated the phylogeny, that is, a pure-birth process with no lineage extinction (Matzke, 2014).
The "extinction" referred to these models is actually the process of range contraction or local
extirpation, which is modeled using the parameter *e*, but *e* is typically dramatically
underestimated (Ree & Smith, 2008; Matzke, 2014).

To test Prediction 1 using model-based inference, each base model was run four times, with or 289 without null ranges in the state space (i.e., BioGeoBEARS' default models vs. "*" models sensu 290 Massana et al. (2015)), and with or without the jump dispersal/founder-event speciation 291 292 parameter i (Matzke, 2014). It is expected that such modifications have a strong impact on 293 inference of dispersal events along branches (i.e., by disallowing null ranges, the anagenetic 294 range contraction rate e can have a much higher estimate, thus raising the probability that a 295 daughter lineage has a different range than the parental lineage) and at cladogenetic events (by adding the parameter *i*). To find the best-supported model given the data, we calculated the 296 297 sample-size corrected Akaike weight AICc (Wagenmakers & Farrell, 2004) for all twelve 298 biogeographical models. We carried out two separate historical biogeography analyses, one at the 299 genus level and the second only using the Indo-Pacific clade.

First, we inferred ancestral ranges using the worldwide phylogeny of *Odontomachus* to better
estimate the ancestral areas at the stem and crown node of the Indo-Pacific clade. Wilson
(Wilson, 1961) hypothesized a SE Asian origin of extant Melanesian *Odontomachus*, but Larabee
et al. (2016) proposed a New World origin. However, it remained unclear in Larabee et al. (2016)
whether trans-Pacific dispersal directly to New Guinea (e.g., via stepping-stone through
archipelagos in Polynesia and eastern Melanesia) or other geographical routes involving SE Asia,

have shaped the extant diversity in the Indo-Pacific. Resolving this issue is important to better 306 307 understand the origin of both Indo-Pacific Odontomachus and the hypothetical taxon cycles. We defined the following geographical areas: 1) Oriental (and Palearctic), encompassing Eurasia, 308 309 continental SE Asia, and the Indo-Malayan region west to Wallace's Line, 2) Wallacea, extending 310 from Wallace's Line to the Lydekker's Line, 3) Philippines, because this region cannot confidently be placed to either side of Wallace's Line, 4) New Guinea, 5) the Bismarck 311 Archipelago and Solomon Islands, 6) Fiji, 7) Australia, 8) New World (Nearctic and Neotropics), 312 and 9) Afrotropics and Malagasy regions. This analysis was time-stratified at 5 Ma, 15 Ma, 25 313 314 Ma and 35 Ma to assign different dispersal rate multipliers and area connectivity through time 315 (Table S1) following paleogeographic models of Earth and the Indo-Pacific (Hall, 2012, 2013). The fossils Odontomachus pseudobauri and O. spinifer were coded as "New World" taxa, and O. 316 paleomyagra was assigned to the "Oriental-Palearctic" area. Biogeographic Stochastic Mapping 317 318 (Dupin et al., 2017) was conducted using the best-fit biogeographical base model (i.e., BayArealike*) to count stochastically-mapped dispersal events out and into New Guinea after the origin 319 320 of the Indo-Pacific clade origin (testing Prediction 2).

Second, we used a trait-dependent dispersal model (Klaus, Matzke et al., in review; for a
simulation test of the model, see Supplemental Information) to allow the influence of a discrete
ecological trait on both anagenetic and cladogenetic dispersal. This was done by assigning rate
multipliers to dispersal depending on the ecological state of a lineage, and by allowing transitions
between ecological states. For this exercise, we used only the Indo-Pacific clade. The code for the
BioGeoBEARS trait-dependent model is available by sourcing the file

327 <u>http://phylo.wdfiles.com/local--files/biogeobears/BioGeoBEARS_traits_v1.R</u> after loading base

BioGeoBEARS. Habitat occurrence for each Indo-Pacific clade species was compiled from the 328 329 literature (Wilson, 1959b, 1959a, 1961; Brown, 1976; Sorger & Zettel, 2011; Sarnat & Economo, 330 2012), public databases (AntWeb, 2017), and our own field records (136 ant colonies each 331 assigned to habitat types) (Table S2). We followed the methodology of recent studies 332 investigating the taxon cycle at a macroevolutionary scale (Economo et al., 2015; Sukumaran et al., 2016) to categorize habitat associations into two states: "forest interior" was assigned to those 333 taxa occurring only at primary interior forests (excluding coastal vegetation), and "open habitat" 334 was assigned to taxa occurring at highly-degraded (secondary) forests, open edge habitats such as 335 336 savannah, and littoral areas. We categorized any species occurring at both primary forest interior 337 and open habitat as the "open habitat" state. Creating a three-state character, adding a state for "both forest and open habitat", is imaginable, but it creates even more free parameters to estimate 338 on a dataset of limited size. Three extra parameters were added in trait-based model variant, 339 340 namely t_{12} and t_{21} (forwards and backwards transition rates between state 1, "open habitat", and state 2, "forest interior"), and m_2 , a multiplier on dispersal probability for the "forest interior" 341 state. The multiplier on dispersal probability for the "open habitat" state, m_1 , was fixed to 1.0, 342 343 whereas m_2 was allowed to range between 0 and 10.0. We conducted extra analyses with m_2 fixed to 1.0, which allowed us to estimate the trait transition rates independently from the 344 345 biogeography parameters, and providing null models to compare to the trait-dependent models. Estimates of transition rates were used to test Prediction 3, and estimates of the multiplier on 346 347 dispersal probability to test Prediction 4.

348 Ancestral state estimation

We used the function ace in the R package ape v4.1 (Paradis et al., 2004) and the functions 349 350 make.simmap and describe.simmap in phytools v0.6-20 (Revell, 2012) to estimate ancestral discrete character states ("forest interior" and "open habitat") on the Indo-Pacific Odontomachus 351 352 clade. Maximum likelihood estimation was conducted under two different models: Equal-rate and 353 All-rates-different. We used two approaches, a continuous-time Markov chain and stochastic character mapping with 1,000 simulations. The ancestral state estimates were used to test 354 Prediction 6, wherein clades having expanded ecological niches (i.e., "open habitat") are the most 355 356 geographically widespread. Such clades are hypothesized to be: 1) the *infandus* group, 2) the ruficeps group, and 3) O. saevissimus. The infandus group consists of taxa at Stage I (O. 357 358 malignus) and Stage III (single-island endemics: O. floresensis, O. angulatus, O. banksi, O. infandus). The ruficeps group consists of taxa at Stage I or II (O. cephalotes, O. ODON015, O. 359 ODON019) and Stage III (endemics to Australia: O. turneri, O. ruficeps). Odontomachus 360 361 saevissimus is in Stage I, and we recognize a new species, sister to O. saevessimus, which is in Stage III (the Halmahera-endemic O. MOLU001). The classification of taxa in Stages belonging 362 363 to taxon cycles follows (Wilson, 1959a, 1961).

364 **Results**

The characteristics of our molecular dataset are presented in Table 3. Overall, the new timecalibrated molecular phylogeny of the genus *Odontomachus* includes 161 ant specimens classified within 36 described species and 16 lineages that most likely represent unrecognized species (Fig. 3; see Fig. S5 for a graphical summary of molecular species delimitation and Fig. S11 for an updated species check-list). This includes 6 times more Indo-Pacific specimens and the identification of 18 more species than in the most recent phylogenetic work on this group 371 (Larabee et al., 2016). We found evidence to support each of the six taxon-cycle predictions
372 outlined in Table 1, suggesting that non-neutral processes have been important in determining the
373 extant species diversity and distribution in the Indo-Pacific.

374 Potential biases in phylogenetic inference

Comparing phylogenies inferred with different data and methods allowed us to determine the 375 robustness of our time-calibrated phylogeny and rule out potential conflicts during the estimation 376 377 of tree topology and branch lengths. First, the single-gene analyses showed no significant discordance in tree topologies as compared to the concatenated multi-locus dataset. Second, 378 similar tree topologies were inferred using different phylogenetic approaches on the same 379 380 concatenated dataset, and the node support tended to agree among analyses (Fig. 2). Critical nodes having low support values were identified in the crown rixosus group (Oriental region) and 381 haematodus group (New World region). These may indicate either a lack of phylogenetic signal 382 383 in our dataset or a poor taxonomic sampling in those regions. Either way, the Indo-Pacific clade, the group we are focusing on in this study, showed relatively good support at critical nodes. 384 385 Third, the tree topology, node support, and divergence time estimates were not significantly 386 different among two fundamentally distinct time-calibration approaches; i.e., the FBD model (Fig. 3) and the calibration-density approach (Fig. S6). Moreover, the estimated root age was not 387 388 driven by the secondary calibration as evidenced by a separate analysis carried out only with priors (Fig. S7). We thus did not detect any of the potential biases in conflicting tree topology and 389 branch length estimations in our time-calibrated phylogeny. 390

391

Historical biogeography of Indo-Pacific Odontomachus: Predictions 1 & 2

To test Predictions 1 and 2 on the ancestral ranges of *Odontomachus*, we used maximum 392 393 likelihood to infer ancestral ranges under different biogeographical models, each making 394 different assumptions about vicariance, within-area speciation, and dispersal. The BayArea-like* 395 model had the best fit to the worldwide phylogeny of *Odontomachus* (AICc weight = 0.52; Table 396 4). This was expected because there were a significant number of within-area cladogenetic events when the defined geographical areas were large and coarse (e.g., the New World; see Figs 3 and 397 S8). On the other hand, when using only the Indo-Pacific clade and a biogeographical model that 398 incorporates habitat-dependent dispersal, the DEC+ $t_{12}+t_{21}+m_2$ model was preferred (AICc weight 399 400 = 0.74; Table 5). For Prediction 1, we found that 1) neither vicariance-heavy models nor models 401 including founder-event speciation were the best fit (i.e., DIVA-like models and models including parameter *j* had lower AICc weights than the best-fit models), and 2) anagenetic range 402 403 expansions lasted long enough on the phylogeny to be detectable by the biogeographical models. 404 For Prediction 2, based on biogeographic histories sampled with Biogeographic Stochastic Mapping (Appendix S2), we showed that dispersal events into New Guinea by taxa that have 405 passed through the taxon cycle (i.e., habitat preference for open environments) are significantly 406 407 younger than dispersal out of New Guinea (t-test p-value = 5.6e–8).

408 Trait-dependent dispersal of Indo-Pacific Odontomachus: Predictions 3 & 4

To test Predictions 3 and 4, we used a trait-dependent dispersal model that allows quantitative estimation of the influence of trait states on macroevolutionary dispersal rates in a single analysis. The best-fit model was the trait-dependent dispersal model DEC+ $t_{12}+t_{21}+m_2$, which allows different transition rates between preference for open habitat and forest interior, and a rate multiplier on dispersal influenced by the "forest interior" state (AICc weight = 0.66; Table 5). Related to Prediction 3, we showed that under the best-fit model transition rate from open habitat to forest interiors ($t_{12} = 0.07$) is at least 3 orders of magnitude higher than the transition from forest interior to open environments ($t_{21} = 0.0001$) (Table 5). For Prediction 4, we showed that there is a significant support for m_2 being different from $m_1 = 1.0$, and in fact the estimate of m_2 was close to 0. This means that habitat preference for forest interior has a dramatic, restrictive influence on macroevolutionary dispersal.

420 Ancestral state estimation of habitat preference: Predictions 5 & 6

To test Predictions 5 and 6 on potential correlations between ecological states related to habitat 421 preference and expanding clades, and their ages, we estimated the ancestral character states using 422 423 maximum likelihood and stochastic character mapping approaches. When modeling the habitat trait alone, the model assuming equal transition rates could not be rejected (p-value > 0.05; the 424 all-rates-different model did not reject the simpler equal-rates model), but the ancestral state 425 estimates overall agree regardless of the chosen model (Figs 4 and S10). For Prediction 5, we 426 found that range-expanding taxa as defined here based on Wilson (1959a, 1961), have significant 427 428 probabilities for a character state shift toward open habitats (Fig. 4), and that Stage I taxa were 429 detected throughout the late Miocene (crown ruficeps group), Pliocene (O. saevissimus, crown infandus group), and Pleistocene (O. malignus). For Prediction 6, we found that taxa with habitat 430 431 preference for open environments, specifically the crown node of *infandus* and *ruficeps* groups, can potentially give rise to species that further re-initiate the taxon cycle (such as O. malignus, O. 432 ODON019) and island endemics such as O. floresensis (Flores), O. angulatus (Fiji), O. banksi 433 and O. infandus (the Philippines). 434

435 Discussion

436

437 consistent with six phylogenetic predictions made by the taxon-cycle model (Table 1). Comparisons among biogeographical models suggest that range expansions and broad ancestral 438 439 geographical distributions lasted long enough to be captured by the biogeographical models. For example, the estimation of within-area "subset" speciation (parameter s in BioGeoBEARS; see 440 Fig. 1) better fit the predicted pattern of widespread ancestral taxa. Moreover, trait-dependent 441 dispersal played an important role in the biogeography of *Odontomachus* in the Indo-Pacific. The 442 443 habitat-dependent dispersal model estimated transition rates from open habitats to forest interior higher than from forest interior to open habitats. Biological interactions, such as competition 444 445 (Wilson, 1961) and parasitism (Ricklefs & Cox, 1972), are hypothesized to drive such a pattern, and explain the transition from Stage I to Stage II and III and the rise of single-island endemics. 446 Furthermore, lineages having preference for forest interior experience drastically reduced 447 448 probabilities of macroevolutionary dispersal compared to lineages having preference for open habitats (Table 5). Overall, our results provide evidence of the important role of non-neutral 449 processes, such as trait-modulated dispersal, on the biogeography of this clade. 450

By using a multi-locus phylogeny of the ant genus *Odontomachus*, we found support for patterns

451 Dispersal events into the Indo-Pacific

Melanesian *Odontomachus* and the hypothetical taxon cycles in the Indo-Pacific originated from
a New World lineage and not from SE Asian rainforest ancestors as previously proposed (Wilson,
1959a, 1961). The initial dispersal event into Melanesia took place in the early Miocene, most
likely as a direct long-distance dispersal event across the Pacific Ocean (see Biogeographic

456	Stochastic Mapping in Appendix S2). Dispersal through Beringia during the mid-Miocene
457	warming followed by extinction of high-latitude taxa cannot be ruled out though, because the
458	biogeographical models used here neglect the process of lineage extinction. However, the early-
459	Pliocene origin of the species O. simillimus (New World haematodus group), currently found in
460	littoral and degraded environments across the Indo-Pacific, possibly reinforces the hypothesis
461	that sporadic trans-Pacific dispersal has contributed to the present-day assemblage of Melanesian
462	fauna. A Gondwanan vicariant origin of the Indo-Pacific and New World clades, which would
463	have left a similar biogeographical pattern, is ruled out by evidence from both fossil (Barden,
464	2017) and molecular divergence times (see also Schmidt 2013; Moreau & Bell 2013). Dispersal
465	from South America to Australia via Antarctica (see Boudinot et al., 2016) is also ruled out given
466	our biogeographical inferences; Australian Odontomachus derive from Melanesian lineages, and
467	not the other way around which would be expected if the ants arrived first to Australia (Fig. 3).
468	There is a growing body of evidence supporting trans-Pacific long-distance dispersal in other
469	insect taxa (e.g., Keppel et al., 2009; Michalak et al., 2010; Gillespie et al., 2012; Birch &
470	Keeley, 2013). Although there are no estimates of the flight dispersal capabilities of
471	Odontomachus, indirect east-to-west dispersal by means of rafts, wood, or debris may have been
472	possible along the South Equatorial Current (Gillespie et al., 2012). Human-mediated dispersal
473	from the New World to the Pacific islands is only likely in the case of Odontomachus ruginodis
474	which we report for the first time in Guam and in New Caledonia (Appendix S1) and given their
475	small genetic divergences among populations (Fig. 2). That O. ruginodis has adapted to urban
476	areas, including docks in the Nearctic region (MacGown et al., 2014), might explain its range
477	expansion aided by commerce. Overall, given the global biogeographic history of the genus

Odontomachus, we hypothesize that trap-jaw ants in general have had the ability to disperse
across remote archipelagos, even along oceanic basins, throughout their evolutionary history.
However, despite their high dispersal ability, we infer that only taxa exploiting resources in open
and disturbed habitats have been successful in colonizing remote areas. This finding departs from
neutral expectations where the observed species distributions are influenced mostly by geography
(e.g., distance and area).

484 The Taxon Cycle disentangled

The species O. simillimus has been considered the prototypical Stage I species, initially 485 expanding from New Guinea toward the Pacific (Wilson, 1961). However, we found a New 486 World origin for this species, and, given the large phylogenetic divergence between it and closely 487 related taxa in the New World, this odd pattern was not likely mediated by human activity. 488 Population genetic data also suggest that early-divergent populations of O. simillimus are found 489 490 in Fiji, suggesting a westward colonization route of the species in Melanesia (Janda et al., 2014). 491 Although O. simillimus did not originate in New Guinea as previously hypothesized (Wilson, 492 1961), other aspects of its biology are consistent with its Stage I status, including wide 493 distribution across biogeographical barriers linked to adaptations to occupy broad habitat types 494 (Economo et al., 2015). The species occurs in disturbed habitats, gardens, forest edges and 495 coastal areas in New Guinea, but in undisturbed primary rainforest on small to medium sized Indo-Pacific islands (e.g., Guadalcanal) (Wilson 1959a; b; Brown 1976; M. Janda & D.M. Sorger 496 pers obs.). Notably, O. simillimus is not established in Madagascar, perhaps due to the presence 497 of the ecologically similar O. troglodytes, also a member of the haematodus group (Fisher & 498 499 Smith, 2008; Larabee et al., 2016). This is in accordance with Wilson's replacement hypothesis,

whereby Stage I taxa might also colonize large source islands where no other ecologically similarspecies exists.

502 Contrary to Wilson's view of SE Asia and New Guinea as the main sources of Indo-Pacific Stage I taxa, our results suggest that the New World has also been a source region of Indo-Pacific taxa. 503 This pattern suggests a more important role of eastern Melanesia, namely Fiji and the Solomon 504 505 Islands, in macroevolutionary dispersal of Stage I taxa in the Indo-Pacific. These areas might have acted as a hub area for New World taxa entering the Indo-Pacific region; a hypothesis that 506 has been put forward in ants (Clouse et al., 2015) and other arthropod groups (Sharma & Giribet, 507 508 2012; Clouse et al., 2017). However, our interpretation does not challenge the importance of SE 509 Asia as a source region for Indo-Pacific taxa. Instead, we propose that an eastern gateway to the Indo-Pacific since the past 15 Ma has also shaped the region's biodiversity. 510

511 Neutral and non-neutral models fitting the biogeography of Odontomachus

512 Our coarse- and fine-scale biogeographic analyses were best fit by different base models, 513 BayArea-like^{*} and DEC+ $t_{12}+t_{21}+m_2$. There is no contradiction between these results as these 514 analyses were done with different phylogenies, and at different temporal and geographical scales. 515 We expect that the finer-scale analysis gives a better approximation to the patterns predicted by 516 the taxon-cycle model, because it includes trait-dependent dispersal parameters, and its focus on 517 the Indo-Pacific region allowed the modeling of range-expansion between geographic areas. Unlike many previous studies of island clades, our statistical model comparison did not indicate 518 improved fit for +i variants. Several observed and inferred ancestral ranges in this study are 519 520 widespread (i.e., occupying multiple areas) suggesting that anagenetic range-expansion dispersal

is common in this group. As predicted by the taxon cycle, *Odontomachus* Stage I taxa have been 521 522 widespread across archipelagos for a considerable time, thus the range-expansion process 523 (modeled by DEC's d parameter) is detectable by biogeographical models. Furthermore, the DEC 524 model allows for "within-area subset speciation" (see Fig. 1), which is consistent with the taxon 525 cycle, in that: 1) speciation begins with peripheral population differentiation of a widespread species, therefore the Stage I in a phylogeny would be represented by ancestral nodes (common 526 ancestors) having widespread distributions, and 2) after a cladogenetic event (i.e., the 527 "instantaneous" speciation on a phylogeny), one of the two daughter lineages would inherit the 528 529 widespread range of the parent node, and the second daughter lineage would inherit a peripheral 530 area, thus entering the process as a single-island endemic.

531 Under a neutral scenario, geographical distances and stochastic dispersal would have shaped the extant distribution of Odontomachus. However, we found little support for the role of 532 geographical distance as a primary regulator of *Odontomachus* macroevolutionary dispersal and 533 534 distribution. Moreover, we found a strong distributional delimitation at Wallace's Line, which is 535 in partial agreement with the hypothesis of biotic interactions (e.g., competition among closely related taxa) influencing geographical distribution (Wilson, 1961); the Oriental rixosus group 536 537 dominates on the western side of the Wallace's Line whereas the Indo-Pacific clade dominates on the eastern side (Sorger & Zettel, 2011; Satria et al., 2015). In other insects, however, it has been 538 reported that the Wallace's Line has been highly permeable, given the strong dispersal abilities of 539 540 winged animals (e.g., Balke et al., 2009; Condamine et al., 2013; Müller et al., 2013; Tänzler et al., 2014; Matos-Maraví et al., 2018), and perhaps due to the continual turnover of species (i.e., 541 542 immigration rate minus extinction rate) across geographically close islands, as expected by

equilibrium theory (Gillespie & Roderick, 2002). For *Odontomachus*, and consistent with
Wilson's narrative, we suggest that ecology (e.g., habitat preference and interspecific interactions
such as competition) has been stronger at delimiting faunas and influencing dispersal than
geography alone at Wallace's Line. In fact, the only two observed *Odontomachus* species that
have spread and established across the *ca*. 35 km distance of the Wallace's Line, *O. simillimus*(New World origin) and *O. malignus* (Melanesian origin), were possibly aided by their ecological
preference for open and disturbed habitats.

Furthermore, under neutral assumptions we expect that species and populations are governed by 550 551 the same evolutionary and migration rates (Missa et al., 2016). However, the geographical spread 552 of *Odontomachus* species have likely been determined by their occupancy of open and disturbed 553 habitats as shown in the trait-dependent dispersal analyses and in studies using other Indo-Pacific ants (Economo & Sarnat, 2012; Economo et al., 2015; Janda et al., 2016; Matos-Maraví et al., 554 2018). As an exemplary case in *Odontomachus*, the crown node of the *infandus* group has likely 555 556 been at Stage I in the late Miocene given the ancestral state inference of habitat preferences, and 557 was followed by range contraction and speciation giving rise to single-island endemics with narrower habitat preferences in Fiji (O. angulatus) (Sarnat & Economo, 2012), Flores (O. 558 559 *floresensis*), and in the Philippines (Sorger & Zettel, 2011). Recent adaptation to exploit 560 resources in coastal habitats occurred in the late Pliocene/early Pleistocene, which favored entering the Stage I by O. malignus (Wilson, 1959b; Brown, 1976; Olsen, 2009). Such a state 561 562 transition likely favored the later expansion of the species across the entire Indo-Pacific, 563 colonizing New Guinea wherein O. malignus only occupies intertidal zones. This biogeographic 564 pattern was expected under the taxon-cycle model, where younger colonizers of "source areas"

are necessarily in Stage I. Although this pattern somewhat resembles the concept of priority 565 566 effects, whereby the ability of species to colonize an island is influenced by preceding colonizers, the priority effect is considered to be stochastic at a macroevolutionary level. This would result in 567 disparate diversity dynamics over time and potentially deviating island communities from 568 569 equilibrium (Lim & Marshall, 2017). But as observed in Odontomachus, the role of priority effects at the macroevolutionary scale may be stronger on species with habitat preference for 570 forest interior and weaker for geographically-expanding taxa occupying open and disturbed 571 habitats. 572

573 However, it is still premature to draw generalizations about the importance of the taxon cycles for insular biogeography. Alternatives to the taxon cycle have been proposed (Pregill & Olson, 1981; 574 Liebherr & Hajek, 1990; Losos, 1992), but its rejection often relies on qualitative approaches. 575 576 Future studies, including meta-analyses using other animal and plant insular clades, need to address additional ecological predictions, such as species abundances of island endemics and 577 578 trait-dependent speciation. For example, species abundances might be a predictor of non-uniform 579 dispersal among species at large regional scales (Shmida & Wilson, 1985), which can be modeled by recent quantitative methods (Rosindell & Phillimore, 2011). However, the most critical 580 581 limitation currently is the lack of records from field expeditions/experiments. In this study, for example, we were unable to recover data on species abundances of Indo-Pacific lineages from 582 either literature or databases. 583

584 Our proposed framework is a useful attempt to study the relative importance of non-neutral and 585 neutral processes in a single biogeographical analysis, but it has several limitations. First, the 586 ancestral state inference for geography and trait evolution does not take into account missing

587 lineages including extinct and unsampled described species. Extinction erases a clade's history 588 (Marshall, 2017) and assuming a Yule (pure-birth) process may mislead if extinction is highly 589 non-random by trait or geographic range. Nevertheless, there is no evidence of mass extinction 590 events in *Odontomachus*, our taxonomic sampling is comprehensive (*ca*. 70% of Indo-Pacific in 591 the phylogeny), and simulation tests indicate that moderate amounts of missing lineages do not 592 severely bias inference (Matzke, 2014).

Second, combining a trait transition matrix with the standard biogeographic transition matrix 593 creates computational challenges. Adding a single binary trait doubles number of states, and 594 595 quadruples the size of the rate matrix, substantially slowing the calculation of the likelihood -a596 calculation which must be repeated many times in a ML analysis. Adding a 3-state trait would 597 increase the size of the transition matrix by 9 times, and adds six transition-rate parameters to the model. Without much larger datasets, this impedes the evaluation of other habitat preferences. 598 For example, in our study we did not explicitly model the influence of the generalist state in 599 600 dispersal (i.e., creating a third state for taxa occupying both forest interiors and open habitats), 601 nor have we attempted to measure the effects of littoral habitat vs. noncoastal open/disturbed environments, despite the fact that coastal taxa might have even greater probabilities to dispersal 602 (Wilson, 1961). 603

604 Conclusion

Non-neutral processes and non-equilibrium dynamics have long been postulated to be important
in island biogeography. The recent developments of quantitative and integrative models that
explicitly incorporate non-neutral processes allow the testing of long-standing, non-equilibrium

hypotheses. In this study, we evaluated six phylogenetic predictions made by the taxon-cycle 608 609 model. We inferred ancestral habitat preferences and investigated the impact of habitat preference on macroevolutionary dispersal. We found strong support for each of the predictions made by the 610 611 taxon cycle, suggesting that non-neutral processes, such as trait-dependent dispersal, have indeed 612 been important for the assemblage and distribution of observed species in the Indo-Pacific. 613 Future studies incorporating other ecological predictions, such as island-endemic abundances over time and trait-dependent diversification, will further clarify whether the taxon cycle is valid 614 and a common observed pattern. Moreover, the predictions outlined in this study can be used in 615 616 future work, including meta-analyses of island communities, to weight the relative contribution 617 of neutral and non-neutral processes in the biogeography and diversification of insular landscapes. 618

619 Acknowledgments

620 We are indebted to the staff of the New Guinea Binatang Research Center and to V. Novotny and S.E. Miller for assistance with our research projects, and to the Papua New Guinea Department of 621 622 Environment and Conservation for assistance with research permits. We are grateful to S. Cover 623 and N.E. Pierce for assistance and access to specimens in MCZ, to A. Andersen and B. Hoffman 624 for kindly providing voucher specimens and DNA sequences for Australian species, M. 625 Borovanska for assistance in the lab, and to D. General and G. Alpert for contributing Philippine samples. We thank Rosemary Gillespie, Evan P. Economo, Søren Faurby, and six anonymous 626 reviewers for constructive comments on previous versions of the manuscript. Funding was 627 provided by the Czech Science Foundation (P505/12/2467), the Putnam Expedition Grants 628 (Museum of Comparative Zoology), GAJU (003/2015/P; 152/2016/P), Marie-Curie fellowship 629

630	(MARIPOSAS-704035), CONACYT DICB-2016 No. 282471, and UNAM PAPIIT IN206818.
631	NJM was supported by a NIMBioS fellowship under NSF award #EFJ0832858, and ARC
632	DECRA fellowship DE150101773. FJL was supported by the National Science Foundation
633	(DDIG DEB-1407279) and the Smithsonian Institution (Peter Buck Fellowship). AVS was
634	supported by the National Geographic Society's Explorers Grant. WCW was supported by funds
635	from the American Museum of Natural History and the US Army Research Laboratory and the
636	US Army Research Office under contract/grant number W911NF-05-1-0271. Computational
637	resources were provided by the MetaCentrum under the program LM2010005 and the CERIT-SC
638	under the program Centre CERIT Scientific Cloud, part of the Operational Program Research and
639	Development for Innovations, Reg. no. CZ.1.05/3.2.00/08.0144.

641 **References**

- de Andrade, M.L. (1994) Fossil Odontomachiti ants from the Dominican Republic (Amber
 Collection Stuttgart: Hymenoptera, Formicidae. VII: Odontomachiti). *Stuttgarter Beiträge zur Naturkunde Serie B (Geologie und Paläontologie)*, **199**, 1–28.
- AntWeb (2017) AntWeb. Available from http://www.antweb.org. Accessed December 2017. *Available from http://www.antweb.org. Accessed December 2017*, .
- Balke, M., Ribera, I., Hendrich, L., Miller, M.A., Sagata, K., Posman, A., Vogler, A.P., & Meier,
 R. (2009) New Guinea highland origin of a widespread arthropod supertramp. *Proceedings of the Royal Society B: Biological Sciences*, 276, 2359–2367.
- Barden, P. (2017) Fossil ants (Hymenoptera: Formicidae): ancient diversity and the rise of
 modern lineages. *Myrmecological News*, 24, 1–30.
- Baroni Urbani, C. (1980) Anochetus corayi n. sp., the first fossil Odontomachiti ant. Stuttgarter
 Beiträge zur Naturkunde Serie B (Geologie und Paläontologie), 55, 1–6.
- Birch, J.L. & Keeley, S.C. (2013) Dispersal pathways across the Pacific: the historical
 biogeography of *Astelia* s.l. (Asteliaceae, Asparagales). *Journal of Biogeography*, 40, 1914–
 1927.
- Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., Suchard, M.A., Rambaut,
 A., & Drummond, A.J. (2014) BEAST 2: a software platform for Bayesian evolutionary
 analysis. *PLoS Computational Biology*, **10**, e1003537.
- Boudinot, B.E., Probst, R.S., Brandão, C.R.F., Feitosa, R.M., & Ward, P.S. (2016) Out of the
 Neotropics: newly discovered relictual species sheds light on the biogeographical history of
 spider ants (*Leptomyrmex*, Dolichoderinae, Formicidae). *Systematic Entomology*, 41, 658–
 663 671.
- Brown, W.L.J. (1976) Contributions toward a reclassification of the Formicidae. Part VI.
 Ponerinae, tribe Ponerini, subtribe Odontomachiti. Section A. Introduction, subtribal
 characters. Genus *Odontomachus. Studia Entomologica*, **19**, 67–171.
- Burnham, K.P. & Anderson, D.R. (2002) Model selection and multimodel inference: a practical
 information-theoretic approach. Springer, New York.
- Clouse, R.M., Branstetter, M.G., Buenavente, P., Crowley, L.M., Czekanski-Moir, J., General,
 D.E.M., Giribet, G., Harvey, M.S., Janies, D.A., Mohagan, A.B., Mohagan, D.P., Sharma,
 P.P., & Wheeler, W.C. (2017) First global molecular phylogeny and biogeographical
 analysis of two arachnid orders (Schizomida and Uropygi) supports a tropical Pangean
 origin and mid-Cretaceous diversification. *Journal of Biogeography*, 44, 2660–2672.
- Clouse, R.M., Janda, M., Blanchard, B., Sharma, P., Hoffmann, B.D., Andersen, A.N.,
 Czekanski-Moir, J.E., Rabeling, C., Wilson, E.O., Economo, E.P., Sarnat, E.M., General,
 D.M., Alpert, G.D., & Wheeler, W.C. (2015) Molecular phylogeny of Indo-Pacific carpenter

- ants (Hymenoptera: Formicidae, *Camponotus*) reveals waves of dispersal and colonization
 from diverse source areas. *Cladistics*, **31**, 424–437.
- Condamine, F.L., Toussaint, E.F.A., Cotton, A.M., Genson, G.S., Sperling, F.A.H., & Kergoat,
 G.J. (2013) Fine-scale biogeographical and temporal diversification processes of peacock
 swallowtails (*Papilio* subgenus *Achillides*) in the Indo-Australian Archipelago. *Cladistics*,
 29, 88–111.
- Cook, B.D., Pringle, C.M., & Hughes, J.M. (2008) Molecular evidence for sequential
 colonization and taxon cycling in freshwater decapod shrimps on a Caribbean island.
 Molecular Ecology, 17, 1066–1075.
- Darriba, D., Taboada, G.L., Doallo, R., & Posada, D. (2012) jModelTest 2: more models, new
 heuristics and parallel computing. *Nature Methods*, 9, 772.
- Dexter, K.G. (2010) The influence of dispersal on macroecological patterns of Lesser Antillean
 birds. *Journal of Biogeography*, 37, 2137–2147.
- Drummond, A.J., Ho, S.Y.W., Phillips, M.J., & Rambaut, A. (2006) Relaxed phylogenetics and
 dating with confidence. *PLoS Biology*, 4, e88.
- Dupin, J., Matzke, N.J., Särkinen, T., Knapp, S., Olmstead, R.G., Bohs, L., & Smith, S.D. (2017)
 Bayesian estimation of the global biogeographical history of the Solanaceae. *Journal of Biogeography*, 44, 887–899.
- Economo, E.P. & Sarnat, E.M. (2012) Revisiting the ants of Melanesia and the taxon cycle:
 historical and human-mediated invasions of a tropical archipelago. *The American Naturalist*, **180**, E1–E16.
- Economo, E.P., Sarnat, E.M., Janda, M., Clouse, R., Klimov, P.B., Fischer, G., Blanchard, B.D.,
 Ramirez, L.N., Andersen, A.N., Berman, M., Guénard, B., Lucky, A., Rabeling, C., Wilson,
 E.O., & Knowles, L.L. (2015) Breaking out of biogeographical modules: range expansion
 and taxon cycles in the hyperdiverse ant genus *Pheidole*. *Journal of Biogeography*, 42,
 2289–2301.
- Fisher, B.L. & Smith, M.A. (2008) A revision of Malagasy species of *Anochetus* Mayr and
 Odontomachus Latreille (Hymenoptera: Formicidae). *PLoS ONE*, 3, e1787.
- Gillespie, R.G., Baldwin, B.G., Waters, J.M., Fraser, C.I., Nikula, R., & Roderick, G.K. (2012)
 Long-distance dispersal: a framework for hypothesis testing. *Trends in Ecology and Evolution*, 27, 47–56.
- Gillespie, R.G. & Roderick, G.K. (2002) Arthropods on islands: colonization, speciation, and
 conservation. *Annual Review of Entomology*, 47, 595–632.
- Goldberg, E.E., Lancaster, L.T., & Ree, R.H. (2011) Phylogenetic inference of reciprocal effects
 between geographic range evolution and diversification. *Systematic Biology*, 60, 451–465.

- Goloboff, P.A. (1999) Analyzing large data sets in reasonable times: solutions for composite
 optima. *Cladistics*, 15, 415–428.
- Goloboff, P.A., Farris, J.S., & Nixon, K.C. (2008) TNT, a free program for phylogenetic analysis.
 Cladistics, 24, 774–786.
- Grimaldi, D. & Engel, M.S. (2005) *Evolution of the Insects*. Cambridge University Press, New
 York, USA.
- Hall, R. (2012) Late Jurassic-Cenozoic reconstructions of the Indonesian region and the Indian
 Ocean. *Tectonophysics*, 570–571, 1–41.
- Hall, R. (2013) The palaeogeography of Sundaland and Wallacea since the Late Jurassic. *Journal of Limnology*, **72**, 1–17.
- Heath, T.A., Huelsenbeck, J.P., & Stadler, T. (2014) The fossilized birth–death process for
 coherent calibration of divergence-time estimates. *Proceedings of the National Academy of Sciences of the United States of America*, 111, E2957–E2966.
- Huelsenbeck, J.P., Larget, B., & Alfaro, M.E. (2004) Bayesian phylogenetic model selection
 using reversible jump Markov chain Monte Carlo. *Molecular Biology and Evolution*, 21,
 1123–1133.
- Iturralde-Vinent, M.A. (2001) Geology of the amber-bearing deposits of the Greater Antilles.
 Caribbean Journal of Science, 37, 141–167.
- Iturralde-Vinent, M.A. & MacPhee, R.D.E. (1996) Age and paleogeographical origin of
 Dominican amber. *Science*, 273, 1850–1852.
- Janda, M., Matos-Maraví, P., Borovanska, M., Zima, J.J., Youngerman, E., & Pierce, N.E. (2016)
 Phylogeny and population genetic structure of the ant genus *Acropyga* (Hymenoptera:
 Formicidae) in Papua New Guinea. *Invertebrate Systematics*, **30**, 28–40.
- Janda, M., Zima, J., Borovanska, M., & Matos-Maraví, P. (2014) Diversification and dispersal of
 Australasian ants, from populations to species. 17th Congress of the International Union for
 the Study of Social Insects (IUSSI). 13 18 July. Cairns, Australia. URI:
 http://hdl.handle.net/2123/11108.
- Janicki, J., Narula, N., Ziegler, M., Guénard, B., & Economo, E.P. (2016) Visualizing and
 interacting with large-volume biodiversity data using client-server web-mapping
 applications: The design and implementation of antmaps.org. *Ecological Informatics*, 32,
 185–193.
- Jønsson, K.A., Irestedt, M., Christidis, L., Clegg, S.M., Holt, B.G., & Fjeldså, J. (2014) Evidence
 of taxon cycles in an Indo-Pacific passerine bird radiation (Aves: Pachycephala).
 Proceedings of the Royal Society B: Biological Sciences, 281, 20131727.
- 746 Keppel, G., Lowe, A.J., & Possingham, H.P. (2009) Changing perspectives on the biogeography

- of the tropical South Pacific: influences of dispersal, vicariance and extinction. *Journal of Biogeography*, **36**, 1035–1054.
- Knor, S., Skuhravá, M., Wappler, T., & Prokop, J. (2013) Galls and gall makers on plant leaves
 from the lower Miocene (Burdigalian) of the Czech Republic: Systematic and
 palaeoecological implications. *Review of Palaeobotany and Palynology*, 188, 38–51.
- Kvaček, Z., Böhme, M., Dvořák, Z., Konzalová, M., Mach, K., Prokop, J., & Rajchl, M. (2004)
 Early Miocene freshwater and swamp ecosystems of the Most Basin (northern Bohemia)
 with particular reference to the Bílina Mine section. *Journal of the Czech Geological Society*, 49, 1–40.
- Landis, M.J., Matzke, N.J., Moore, B.R., & Huelsenbeck, J.P. (2013) Bayesian analysis of
 biogeography when the number of areas is large. *Systematic Biology*, 62, 789–804.
- Lanfear, R., Calcott, B., Ho, S.Y.W., & Guindon, S. (2012) PartitionFinder: combined selection
 of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution*, 29, 1695–1701.
- Larabee, F.J., Fisher, B.L., Schmidt, C.A., Matos-Maraví, P., Janda, M., & Suarez, A. V. (2016)
 Molecular phylogenetics and diversification of trap-jaw ants in the genera *Anochetus* and
 Odontomachus (Hymenoptera: Formicidae). *Molecular Phylogenetics and Evolution*, 103,
 143–154.
- Liebherr, J.K. & Hajek, A.E. (1990) A cladistic test of the taxon cycle and taxon pulse
 hypotheses. *Cladistics*, 6, 39–59.
- Lim, J.Y. & Marshall, C.R. (2017) The true tempo of evolutionary radiation and decline revealed
 on the Hawaiian archipelago. *Nature*, 543, 710–713.
- Liu, C., Guénard, B., Garcia, F.H., Yamane, S., Blanchard, B., Yang, D.R., & Economo, E.
 (2015) New records of ant species from Yunnan, China. *ZooKeys*, 477, 17–78.
- Losos, J.B. (1992) A critical comparison of the Taxon-Cycle and Character Displacement models
 for size evolution of *Anolis* lizards in the Lesser Antilles. *Copeia*, 2, 279–288.
- Lucky, A. & Sarnat, E.M. (2010) Biogeography and diversification of the Pacific ant genus
 Lordomyrma Emery. *Journal of Biogeography*, **37**, 624–634.
- MacArthur, R.H. & Wilson, E.O. (1967) *The Theory of Island Biogeography*. Princeton
 University Press, Princeton, NJ.
- MacGown, J.A., Boudinot, B., Deyrup, M., & Sorger, D.M. (2014) A review of the Nearctic
 Odontomachus (Hymenoptera: Formicidae: Ponerinae) with a treatment of the males.
 Zootaxa, 3802, 515–552.
- Marshall, C.R. (2017) Five palaeobiological laws needed to understand the evolution of the living
 biota. *Nature Ecology and Evolution*, 1, 0165.

- Massana, K.A., Beaulieu, J.M., Matzke, N.J., & O'Meara, B.C. (2015) Non-null effects of the
 null range in biogeographic models: Exploring parameter estimation in the DEC model.
 bioRxiv, September, 1–25, doi: 10.1101/026914.
- Matos-Maraví, P., Clouse, R.M., Sarnat, E.M., Economo, E.P., LaPolla, J.S., Borovanska, M.,
 Rabeling, C., Czekanski-Moir, J.E., Latumahina, F., Wilson, E.O., & Janda, M. (2018) An
 ant genus-group (*Prenolepis*) illuminates the drivers of insect diversification in the IndoPacific. *Molecular Phylogenetics and Evolution*, **123**, 16–25.
- Matzke, N.J. (2013a) BioGeoBEARS: BioGeography with Bayesian (and Likelihood)
 Evolutionary Analysis in R Scripts. http://cran.r-project.org/package=BioGeoBEARS.
- Matzke, N.J. (2013b) Probabilistic historical biogeography: new models for founder-event
 speciation, imperfect detection, and fossils allow improved accuracy and model-testing.
 Frontiers of Biogeography, 5, 242–248.
- Matzke, N.J. (2014) Model selection in historical biogeography reveals that founder-event
 speciation is a crucial process in island clades. *Systematic Biology*, 63, 951–970.
- Michalak, I., Zhang, L.B., & Renner, S.S. (2010) Trans-Atlantic, trans-Pacific and trans-Indian
 Ocean dispersal in the small Gondwanan Laurales family Hernandiaceae. *Journal of Biogeography*, 37, 1214–1226.
- Miller, M.A., Pfeiffer, W., & Schwartz, T. (2010) Creating the CIPRES Science Gateway for
 inference of large phylogenetic trees. *Proceedings of the Gateway Computing Environments Workshop (GCE), 14 Nov.*, 1–8.
- Missa, O., Dytham, C., & Morlon, H. (2016) Understanding how biodiversity unfolds through
 time under neutral theory. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **371**, 20150226.
- Moreau, C.S. & Bell, C.D. (2013) Testing the museum versus cradle tropical biological diversity
 hypothesis: phylogeny, diversification, and ancestral biogeographic range evolution of the
 ants. *Evolution*, 67, 2240–2257.
- Müller, C.J., Matos-Maraví, P., & Beheregaray, L.B. (2013) Delving into *Delias* Hübner
 (Lepidoptera: Pieridae): fine-scale biogeography, phylogenetics and systematics of the
 world's largest butterfly genus. *Journal of Biogeography*, **40**, 881–893.
- Nixon, K.C. (1999) The parsimony ratchet, a new method for rapid parsimony analysis.
 Cladistics, 15, 407–414.
- Olsen, A.R. (2009) New record of the marine littoral ant, *Odontomachus malignus* Smith, F.
 1859, in Palau. *The Pan-Pacific Entomologist*, **85**, 11–12.
- Paradis, E., Claude, J., & Strimmer, K. (2004) APE: Analyses of Phylogenetics and Evolution in
 R language. *Bioinformatics*, 20, 289–290.

- Parham, J.F., Donoghue, P.C.J., Bell, C.J., et al. (2012) Best practices for justifying fossil
 calibrations. *Systematic Biology*, 61, 346–359.
- Patiño, J., Whittaker, R.J., Borges, P.A.V., et al. (2017) A roadmap for island biology: 50
 fundamental questions after 50 years of *The Theory of Island Biogeography. Journal of*
- Biogeography, 44, 963–983.
- Pregill, G.K. & Olson, S.L. (1981) Zoogeography of West Indian vertebrates in relation to
 Pleistocene climatic cycles. *Annual Review of Ecology and Systematics*, 12, 75–98.
- R Core Team (2017) R: A language and environment for statistical computing. Version 3.4.2.
 http://www.r-project.org.
- Ree, R.H., Moore, B.R., Webb, C.O., & Donoghue, M.J. (2005) A likelihood framework for
 inferring the evolution of geographic range on phylogenetic trees. *Evolution*, **59**, 2299–2311.
- Ree, R.H. & Smith, S.A. (2008) Maximum likelihood inference of geographic range evolution by
 dispersal, local extinction, and cladogenesis. *Systematic Biology*, 57, 4–14.
- Revell, L.J. (2012) phytools: an R package for phylogenetic comparative biology (and other
 things). *Methods in Ecology and Evolution*, 3, 217–223.
- Ricklefs, R. & Bermingham, E. (2008) The West Indies as a laboratory of biogeography and
 evolution. *Philosophical Transactions of the Royal Society London Series B*, 363, 2393–
 2413.
- Ricklefs, R.E. & Bermingham, E. (2002) The concept of the taxon cycle in biogeography. *Global Ecology and Biogeography*, 11, 353–361.
- Ricklefs, R.E. & Cox, G.W. (1972) Taxon cycles in the West Indian avifauna. *The American Naturalist*, **106**, 195–219.
- Ronquist, F. (1997) Dispersal-vicariance analysis: a new approach to the quantification of
 historical biogeography. *Systematic Biology*, 46, 195–203.
- Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B.,
 Liu, L., Suchard, M.A., & Huelsenbeck, J.P. (2012) MrBayes 3.2: efficient Bayesian
 phylogenetic inference and model choice across a large model space. *Systematic Biology*,
 61, 539–542.
- Rosindell, J. & Phillimore, A.B. (2011) A unified model of island biogeography sheds light on
 the zone of radiation. *Ecology Letters*, 14, 552–560.
- 847 Sarnat, E.M. & Economo, E.P. (2012) *The Ants of Fiji*. University of California Press, Berkeley,
 848 California.
- Satria, R., Kurushima, H., Herwina, H., Yamane, S., & Eguchi, K. (2015) The trap-jaw ant genus
 Odontomachus Latreille (Hymenoptera: Formicidae) from Sumatra, with a new species

- 851 description. *Zootaxa*, **4048**, 1–36.
- Schmidt, C. (2013) Molecular phylogenetics of ponerine ants (Hymenoptera: Formicidae:
 Ponerinae). *Zootaxa*, 3647, 201–250.
- Schmidt, C.A. & Shattuck, S.O. (2014) The higher classification of the ant subfamily Ponerinae
 (Hymenoptera: Formicidae), with a review of ponerine ecology and behavior. *Zootaxa*,
 3817, 1–242.
- Schultz, T.R. & Brady, S.G. (2008) Major evolutionary transitions in ant agriculture. *Proceedings of the National Academy of Sciences of the United States of America*, 105, 5435–5440.
- Sharma, P.P. & Giribet, G. (2012) Out of the Neotropics: Late Cretaceous colonization of
 Australasia by American arthropods. *Proceedings of the Royal Society B: Biological Sciences*, 279, 3501–3509.
- Shmida, A. & Wilson, M. V. (1985) Biological determinants of species diversity. *Journal of Biogeography*, **12**, 1–20.
- Sorger, D.M. & Zettel, H. (2011) On the ants (Hymenoptera: Formicidae) of the Philippine
 Islands: V. The genus *Odontomachus* Latreille, 1804. *Myrmecological News*, 14, 141–163.
- Stamatakis, A. (2014) RAxML version 8: A tool for phylogenetic analysis and post-analysis of
 large phylogenies. *Bioinformatics*, **30**, 1312–1313.
- Sukumaran, J., Economo, E.P., & Knowles, L.L. (2016) Machine learning biogeographic
 processes from biotic patterns: a new trait-dependent dispersal and diversification model
 with model choice by simulation-trained discriminant analysis. *Systematic Biology*, 65, 525–
 545.
- Sukumaran, J. & Knowles, L.L. (2018) Trait-dependent biogeography: (re)integrating biology
 into probabilistic historical biogeographical models. *Trends in Ecology & Evolution*, In
 Press.
- Tänzler, R., Toussaint, E.F.A., Suhardjono, Y.R., Balke, M., & Riedel, A. (2014) Multiple
 transgressions of Wallace's Line explain diversity of flightless *Trigonopterus* weevils on
 Bali. *Proceedings of the Royal Society B: Biological Sciences*, 281, 20132528.
- Valente, L.M., Phillimore, A.B., & Etienne, R.S. (2015) Equilibrium and non-equilibrium
 dynamics simultaneously operate in the Galápagos islands. *Ecology Letters*, 18, 844–852.
- Wagenmakers, E.-J. & Farrell, S. (2004) AIC model selection using Akaike weights. *Psychonomic Bulletin and Review*, **11**, 192–196.
- Wappler, T., Dlussky, G.M., Engel, M.S., Prokop, J., & Knor, S. (2014) A new trap-jaw ant
 species of the genus *Odontomachus* (Hymenoptera: Formicidae: Ponerinae) from the Early
 Miocene (Burdigalian) of the Czech Republic. *Paläontologische Zeitschrift*, 88, 495–502.

- Whittaker, R.J., Fernández-Palacios, J.M., Matthews, T.J., Borregaard, M.K., & Triantis, K.A.
 (2017) Island biogeography: taking the long view of nature's laboratories. *Science*, 357, eaam8326.
- Wilson, E.O. (1959a) Adaptive shift and dispersal in a tropical ant fauna. *Evolution*, **13**, 122–144.
- Wilson, E.O. (1959b) Studies on the ant fauna of Melanesia. V. The tribe Odontomachini.
 Bulletin of the Museum of Comparative Zoology at Harvard College, **120**, 483–510.
- Wilson, E.O. (1961) The nature of the taxon cycle in the Melanesian ant fauna. *The American Naturalist*, **95**, 169–193.
- Yoshimura, M., Onoyama, K., & Ogata, K. (2007) The ants of the genus *Odontomachus* (Insecta:
 Hymenoptera: Formicidae) in Japan. *Species Diversity*, 12, 89–112.
- 895

897 Data Accessibility

- DNA sequences: Genbank accessions KU145821-KU146453; BOLD: under ASPNA project.
- Datasets and time-calibrated phylogenies: TreeBase study ID 202312.
- Input files for phylogenetic analyses and R code for biogeographical analyses; sampling
 localities: Dryad Digital Repository, doi: https://doi.org/10.5061/dryad.5542pr8.
- 902 Specimens deposited: Harvard Museum of Comparative Zoology (MCZ), Smithsonian National
- 903 Museum of Natural History (USNMENT), CSIRO Tropical Ecosystems Research Centre,
- 904 Darwin, Australia (TERC), Institute of Entomology, Czech Academy of Sciences (EntU-CAS).

905 Author Contributions

- 906 PMM and MJ designed the study; PMM, FJL, RMC, WCW, DMS, AVS and MJ collected
- specimens; PMM and FJL performed laboratory work; PMM and NJM conducted analyses;
- 908 PMM wrote the first draft of the manuscript and all co-authors wrote the final version of the 909 article.

910 Supporting Information

- 911 Additional Supporting Information may be found in the online version of this article:
- 912 Fig. S1. Sampling localities plotted in a map of the Indo-Pacific.
- Fig. S2. Single gene phylogenies inferred in MrBayes v.3.2.3.
- 914 Fig. S3. Molecular species delimitation by mPTP.
- Fig. S4. Molecular species delimitation by BP&P v.3.1.
- Fig. S5. Molecular species delimitation between mPTP and BP&P v3.1.
- 917 Fig. S6. Maximum clade credibility ime-calibrated phylogeny using a calibration-density
- 918 approach in BEAST v.2.3.1.
- Fig. S7. Root age's posterior distributions compared to the distribution using only the priors inBEAST v.2.3.1.
- Fig. S8. Inferred ancestral ranges using the BayArea-like* model on the worldwide phylogeny of*Odontomachus*.
- 923 Fig. S9. Inferred ancestral ranges using the DEC+ $t_{12}+t_{21}+m_2$ model on the Indo-Pacific clade.
- Fig. S10. Ancetral habitat preference estimates using the R packages *ape* and *phytools*.
- 925 Fig. S11. *Odontomachus* species check-list in a phylogenetic context.
- 926 Table S1. Matrices with allowed areas and manual dispersal multipliers used for the time-

- 927 stratified analyses in BioGeoBEARS v.0.2.1.
- Table S2. Number of own field records and literature data used to categorize habitat states perspecies in the habitat-dependent dispersal analyses.
- Appendix S1. Voucher information, habitat association for Indo-Pacific clade species, and DNA
 sequences accession numbers in GenBank and BOLD.
- 932 Appendix S2. Biogeographical stochastic mappings using the best-fit biogeographical model for
- 933 the worldwide phylogeny of *Odontomachus* (BayArea-like*).

935 Main manuscript's figures and tables

936 Fig1. Macroevolutionary and biogeographical patterns expected under the taxon-cycle model:

937 Stage I, anagenetic (population level) range expansion aided by a shift in habitat preference to

- 938 open or disturbed environments; Stage II, differentiation among populations, leaving a patchy
- 939 distribution due to anagenetic range contraction and within-area "subset" cladogenesis, wherein
- one of the two daughter lineages inherits only a smaller area within a widespread parental
- species; Stage III, single-island endemics evolve mainly through the cladogenetic "subset"
- 942 process and re-entering Stage I, linked again to recent ecological broadening.





- Fig. 2. Phylogeny of *Odontomachus* plotted to the 50% majority rule consensus tree from
- 945 MrBayes using the concatenated multi-locus. Support values are depicted as colored stars in the 946 following order: MrBayes posterior probability (PP), RAxML bootstraps (BP), and TNT
- following order: MrBayes posterior probability (PP), RAxML bootstraps (BP), and TNT
 bootstraps (BP). Black stars represent high support (PP > 0.95 and BP >90), blue stars represent
- moderate support (PP 0.90-0.94 and to BP 70-89), and red stars indicate low support (PP < 0.89
- and to BS < 69). Lower support are indicated with dashes. The Indo-Pacific species groups

defined by Brown (1976) based on morphology were recovered with moderate to high support, except for the *infandus* group, which originally included the *papuanus* group.



Fig. 3. Maximum Clade Credibility (MCC) tree inferred in BEAST using the Fossilized-Birth-953 954 Death (FBD) model. Posterior probability color code follows Fig. 2. The 95% dating confidence intervals are depicted as blue horizontal bars. Extant distributions of species are depicted as 955 colored squares following the legend and map in the figure. Inferred ancestral ranges based on 956 the coarse-scale biogeographic analysis (BayArea-like* model) are displayed as color blocks on 957 the nodes, with probabilities of the most likely ranges as black in the associated pie chart. Panels 958 959 A: O. rixosus, B: O. simillimus, C: O. tyrannicus, D: O. malignus; scale bar in red represents 1 960 mm.



Fig. 4. Ancestral area inference on the Indo-Pacific clade using a trait-dependent dispersal model implemented in BioGeoBEARS. The tree topology, divergence times, and posterior probabilities are identical as the phylogeny shown in Fig. 3. Extant distributions of species are depicted as colored square(s). Inferred ancestral ranges based on the DEC+ $t_{12}+t_{21}+m_2$ model are displayed on

- 966 main nodes. The probability of the most probable inferred range is indicated by black in the
- 967 associated pie chart, and habitat preference is depicted as colored circles. Ancestral habitat
- 968 preferences were estimated in phytools under the continuous-time Markov chain and stochastic
- 969 character mapping with 1,000 simulations. Inferred nodes with the open habitat state
- 970 (open/disturbed environments; "A.S. O/D") are shown as colored arrows in blue (probability >
- 971 (0.5) and light-blue (probability > 0.25).





- 974 Table 1
- 975 Six phylogenetic predictions expected if a taxon cycle is operating, and alternative predictions if
- a taxon cycle is not operating. These are predictions about what will be inferred from model-
- 977 based inference of ancestral ecological traits (habitat preferences), ancestral geographic ranges, or
- both (in the case of a joint trait-based dispersal, biogeographical model).

Prediction	Observations and/or Phylogeny-based inference	If the Taxon Cycle model is operating in a clade, we expect:	If the Taxon Cycle is not operating
1	Inference of ancestral geographic ranges (standard BioGeoBEARS models)	Range expansions (e.g., from 1 area to 2+areas) will last long enough to be detectable on the branches of a phylogeny. This will be shown by estimated values of parameter <i>d</i> greater than 0 in DEC-type biogeography models, and the commonality of multi-area ranges in Biogeographical Stochastic Mapping.	Dominance of single-area ancestral ranges, and dispersal events dominated by jump dispersal/founder event-speciation controlled by parameter <i>j</i> , thus significantly improving the fit of the model to the dataset; alternatively, a scenario dominated by ancient widespread ranges that break up through vicariance.
2	Timing of dispersal to/from the "source area" (here, New Guinea)	Dispersals out of the "source area" are older than into "source area" by taxa at Stage I	No significant difference in timing of dispersal events into and out of the source area
3	Inferred transition rates between ecological states (forest interior versus open/disturbed habitat)	Transition rates for "open" habitats to forest interiors higher than the opposite, because taxa occupying open/disturbed habitats encounter two "evolutionary fates": 1) extinction, or 2) survival in the forest interior. Such "evolutionary fates" may be controlled by biological interactions such as competition and parasitism	No significant difference in transition rates between habitat preferences
4	Trait-dependent dispersal model	Dispersal occurs entirely or primarily in lineages occupying open/disturbed habitats, such as islands coasts, because they are more prone to colonize other archipelagos that taxa restricted only to forest interiors. Habitat preference for forest interior thus have low or zero dispersal.	No significant different dispersal rates between lineages with distinct ecologies. Dispersal modulated by neutral processes such as geographical distance and area
5	Timing of range expansion and ecological shift events	At different times across clades	At approximately the same times across clades (environment-shift mediated; e.g., Pleistocene sea-level fluctuations)
6	Species ranges and ecology	A few widespread taxa with broad ecological niches, potentially within 1) the <i>infandus</i> group, 2) the <i>ruficeps</i> group, and 3) O. <i>saevissimus</i> . Most taxa single-area endemics	Widespread taxa regardless of ecological preference

980 Table 2

981 Brief description and taxonomic affinities of the five fossil taxa used in the calibration of

982 Odontomachus phylogeny. Odontomachus paleomyagra was found in a coal seam, Most Basin

983 locality, Czech Republic, and the remaining fossil taxa were found in Dominican amber

984 inclusions.

Fossil Taxon	Taxonomic affinity and synapomorphies	Holotype			
Anochetus corayi Baroni Urbani, 1980	<i>A. mayri</i> lineage; Squamiform excise petiole	Gyne; No. Do-834-K-1, State Museum of Natural History, Sttutgart			
Anochetus exstinctus De Andrade, 1994	<i>A. emarginatus</i> lineage; Serially dentate mandibles, petiolar and propodeal spines	Worker; No. Do-5479, State Museum of Natural History, Sttutgart			
Odontomachus paleomyagra Wappler et al., 2014	<i>O. assiniensis-rixosus</i> clade; Mandibles constriction	Gyne; No. ZD0136, Bílina Mine Enterprises, Czechia			
Odontomachus pseudobauri De Andrade, 1994	<i>O. haematodus</i> species group; Metasternal process	Worker; Amber Sample A, Natural History Museum, London			
Odontomachus spinifer De Andrade, 1994	<i>O. haematodus</i> species group; Smooth vertex	Worker; No. Do-2215, State Museum of Natural History, Sttutgart			

985

986 Table 3

987 Molecular composition of the data matrix used in the time-calibrated divergence analyses. Six

loci and 52 specimens representing 52 species were analyzed for testing the phylogenetic

989 predictions of the taxon cycle hypothesis. (Pis) = Number of parsimony-informative sites.

Genes	Specimens	Length (bp)	ength (bp) Variable sites (Pis) Missing data (%		GC content (%)	
285	46 (89%)	889	113 (48)	27.0	63.3	
CAD	39 (75%)	844	212 (130) 41.4		48.6	
EF-1αF1	29 (56%)	359	47 (23)	50.4	62.6	
COI	46 (89%)	659	299 (214)	17.9	24.3	
LWR	46 (89%)	583	184 (108)	16.5	52.3	
wingless	50 (96%)	421	116 (80)	6.3	61.9	
TOTAL	52	3755	971 (603)	26.9	50.8	

990

991 Table 4

992 BioGeoBEARS analyses of biogeography alone (no traits included), at a coarse geographic scale

- 993 (across multiple continents). Twelve models were evaluated, including the DEC, DIVA-like, and
- BayArea-like base models, either allowing or not (* models) null ranges, and either including or
- not an extra parameter *j*. Estimated parameters: *d*: anagenetic range expansion; *e*: anagenetic
- ⁹⁹⁶ range contraction; *j*: cladogenetic founder-event. AICc: Sample-size corrected Akaike
- 997 information criterion, *df*: degrees of freedom. The best-fit biogeographical model to the time-
- 998 calibrated tree of the genus *Odontomachus* was the BayArea-like* (i.e., disallowing null ranges).

Biogeographic models	d	е	j	loglik	AICc	df	Akaike Weights
DEC	0.022	0.006	0.000	-116.640	237.547	2	0.00
DEC + J	0.021	0.005	0.018	-116.170	238.885	3	0.00
DEC*	0.050	0.305	0.000	-107.620	219.507	2	0.15
DEC* + J	0.051	0.305	0.000	-107.620	221.785	3	0.05
BayArea-like	0.025	0.039	0.000	-137.660	279.587	2	0.00
BayArea-like + J	0.019	0.012	0.102	-125.990	258.525	3	0.00
BayArea-like*	0.044	0.268	0.000	-106.390	217.047	2	0.52
BayArea-like* + J	0.045	0.286	0.000	-106.370	219.285	3	0.17
DIVA-like	0.030	0.010	0.000	-126.770	257.807	2	0.00
DIVA-like + J	0.023	0.008	0.069	-123.870	254.285	3	0.00
DIVA-like*	0.053	0.360	0.000	-108.250	220.767	2	0.08
DIVA-like* + J	0.052	0.339	0.000	-108.240	223.025	3	0.03

999

1000 Table 5

1001 BioGeoBEARS analyses of biogeography+trait data (comparing models with trait-independent dispersal and trait-dependent dispersal), at a fine geographic scale (the Indo-Pacific clade, 26 1002 extant tips). Twelve models were evaluated, using the DEC, DIVA-like, and BayArea-like base 1003 models either including or not an extra parameter *i*. Estimated parameters: *d*: anagenetic range 1004 expansion; e: anagenetic range contraction; j: cladogenetic founder-event; t_{12} : transition rate from 1005 1006 open habitat (open/disturbed environments) to forest interior; t_{21} : transition rate from forest 1007 interior to open habitat; m_1 : multiplier on dispersal rate when in "open habitat" state was fixed to 1.0 in all models; m_2 : multiplier on dispersal rate when in "forest interior" state. AICc: Sample-1008 1009 size corrected Akaike information criterion, df: degrees of freedom (number of free parameters). 1010 The best-fit biogeographical model was DEC+ $t_{12}+t_{21}+m_2$.

Biogeographic models	d	e	j	t ₁₂	t ₂₁	m_2	loglik	AICc	df	Akaike Weights
DEC+2rates, $m_2=1$	0.007	0.000	0.000	0.074	0.000	1.000	-67.167	144.238	4	0.02
DEC+J+2rates, $m_2=1$	0.007	0.000	0.006	0.074	0.000	1.000	-66.547	146.094	5	0.01
DEC+2rates+ m_2	0.016	0.000	0.000	0.072	0.000	0.000	-61.883	136.767	5	0.66
DEC+J+2rates+ m_2 , run1	0.014	0.000	0.035	0.074	0.000	0.000	-60.941	138.302	6	0.31
DIVA+2rates, $m_2=1$	0.012	0.000	0.000	0.074	0.000	1.000	-75.615	161.136	4	0.00
DIVA+J+2rates, $m_2=1$	0.009	0.000	0.021	0.074	0.000	1.000	-73.756	160.513	5	0.00
DIVA+2rates+ m_2	0.024	0.000	0.000	0.072	0.000	0.000	-66.810	146.620	5	0.00
DIVA+J+2rates+ m_2 , run1	0.017	0.000	0.069	0.073	0.000	0.000	-64.327	145.076	6	0.01
BAYAREA+2rates, $m_2=1$	0.006	0.076	0.000	0.074	0.000	1.000	-79.442	168.789	4	0.00
BAYAREA+J+2rates, $m_2=1$	0.007	0.000	0.038	0.074	0.000	1.000	-78.110	169.220	5	0.00
BAYAREA+2rates+ m_2	0.013	0.075	0.000	0.072	0.000	0.000	-75.434	163.869	5	0.00
BAYAREA+J+2rates+ m_2 , run2	0.014	0.000	0.131	0.074	0.000	0.000	-68.018	152.456	6	0.00