

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25

DR. JOSEPH DIBATTISTA (Orcid ID : 0000-0002-5696-7574)

DR. JONATHAN CHANG (Orcid ID : 0000-0002-3811-1254)

Article type : Original Research

Article Type: Original Article

Title: Ice ages and butterflyfishes: phylogenomics elucidates the ecological and evolutionary history of reef fishes in an endemism hotspot

Running Head: Evolution of endemic butterflyfish

Joseph D. DiBattista^{a,b,c*}, Michael E. Alfaro^d, Lori Sorenson^d, John H. Choat^e, Jean-Paul A. Hobbs^c, Tane H. Sinclair-Taylor^a, Luiz A. Rocha^f, Jonathan Chang^d, Osmar J. Luiz^g, Peter F. Cowman^h, Matt Friedman^{ij}, and Michael L. Berumen^a

^aRed Sea Research Center, Division of Biological and Environmental Science and Engineering, King Abdullah University of Science and Technology, Thuwal 23955, Saudi Arabia

^bAustralian Museum Research Institute, Australian Museum, 1 William St, Sydney, NSW 2010, Australia

^cSchool of Molecular and Life Sciences, Curtin University, Perth, WA 6845, Australia

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as [doi: 10.1002/ECE3.4566](https://doi.org/10.1002/ECE3.4566)

This article is protected by copyright. All rights reserved

26 ^dDepartment of Ecology and Evolutionary Biology, University of California Los Angeles, Los
27 Angeles, CA 90095, USA

28

29 ^eSchool of Marine and Tropical Biology, James Cook University, Townsville, QLD 4811,
30 Australia

31

32 ^fSection of Ichthyology, California Academy of Sciences, San Francisco, CA 94118, USA

33

34 ^gResearch Institute for the Environment and Livelihoods, Charles Darwin University, Darwin,
35 NT 0810, Australia

36

37 ^hARC Centre of Excellence for Coral Reef Studies, James Cook University, Townsville,
38 Australia

39

40 ⁱDepartment of Earth Sciences, University of Oxford, South Parks Road, Oxford, OX1 3AN, UK

41

42 ^jMuseum of Paleontology and Department of Earth and Environmental Sciences, Ann Arbor, MI
43 48109, USA

44

45

46 *Correspondence: *Australian Museum Research Institute, Australian Museum, 1 William St,*
47 *Sydney, NSW 2010, Australia*

48 Phone: +61 0 478 514 097

49 E-mail: josephdibattista@gmail.com

50

51 **Abstract**

52 For tropical marine species, endemism hotspots occur in peripheral areas furthest from the center
53 of diversity, but the evolutionary processes that lead to their origin remains elusive. We test
54 several hypotheses related to the evolution of peripheral endemics by sequencing ultraconserved
55 element (UCE) loci to produce a genome-scale phylogeny of 47 butterflyfish species (family

56 Chaetodontidae) that includes all shallow-water butterflyfish from the coastal waters of the
57 Arabian Peninsula (i.e. Red Sea to Arabian Gulf) and their close relatives. Bayesian tree building
58 methods produced a well-resolved phylogeny that elucidated the origins of butterflyfishes in this
59 endemism hotspot. We show that UCEs, often used to resolve deep evolutionary relationships,
60 represent an important tool to assess the mechanisms underlying recently diverged taxa. Our
61 analyses indicate that unique environmental conditions in the coastal waters of the Arabian
62 Peninsula probably contributed to the formation of endemic butterflyfishes. Older endemic
63 species are also associated with narrow versus broad depth ranges, suggesting that adaptation to
64 deeper coral reefs in this region occurred only recently (< 1.75 Ma). Even though deep reef
65 environments were drastically reduced during extreme low sea level stands of glacial ages,
66 shallow reefs persisted, and as such there was no evidence supporting mass extirpation of fauna
67 in this region.

68
69
70
71

72 Keywords: biogeographic barriers; *Chaetodon*; coral reef; glaciation events; Pleistocene;
73 ultraconserved elements

74 **Introduction**

75 Explaining the underlying factors responsible for the diversity of species accumulated at
76 endemism hotspots remains a difficult problem in the field of biogeography. Recent research has
77 identified the importance of peripheral regions from tropical oceans in generating and exporting
78 biological diversity, thus intermittently seeding adjacent seas (DiBattista *et al.* 2010; Eble *et al.*
79 2011; Gaither *et al.* 2010, 2011; Malay & Paulay 2010; Skillings *et al.* 2010; Bowen *et al.* 2013;
80 DiBattista *et al.* 2013), however, direct tests of this assumption are rare. Renewed interest in the
81 Red Sea to Arabian Gulf (or Persian Gulf) region provides a new opportunity to explore
82 hypotheses associated with how endemics are formed in peripheral areas, and its potential
83 contribution to the species richness of marine biodiversity hotspots. The Red Sea is a semi-
84 enclosed basin located at the north-western corner of the Indian Ocean and harbors one of the
85 highest levels of endemism for marine organisms (12.9% for fishes, 12.6% for polychaetes, 8.1%

86 for echinoderms, 16.5% for ascidians, and 5.8% for scleractinian corals; DiBattista *et al.* 2016a).
87 The level of endemism among well-characterized groups in the Red Sea, such as the shore fishes,
88 exceeds those of all other peripheral endemic hotspots identified for the Indian Ocean (DiBattista
89 *et al.* 2016a). Although many of these Red Sea endemics extend their distribution into the
90 adjacent Gulf of Aden and Arabian Sea (Kemp 1998; DiBattista *et al.* 2016a,b), it is not clear
91 whether they are paleo-endemics (old lineages restricted due to range contraction), neo-endemics
92 (young lineages at site of origin), or “ecological” endemics (old or young lineages restricted
93 range due to species ecology; see Cowman *et al.* 2017) and where, when, and how this
94 diversification occurred.

95 The Red Sea has a unique geological and paleoclimactic history that may have played a
96 role in its high levels of endemism (see DiBattista *et al.* 2016b for review). In brief, the Red Sea
97 basin was formed by episodes of sea floor spreading 41 to 34 Ma (Girdler & Styles 1974),
98 followed by intermittent connections to the Mediterranean Sea in the north (~14 Ma to 5 Ma;
99 Hubert-Ferrari *et al.* 2003), and a more recent connection to the Gulf of Aden in the south
100 through the Strait of Bab al Mandab (~5 Ma to present; Bailey 2009). The Strait is a narrow
101 channel (29 km) with a shallow sill (137 m) that constitutes the only connection between the Red
102 Sea and the Indian Ocean (Bailey 2009). Water exchange is regulated by Indian Ocean monsoon
103 patterns (Smeed 1997; Raitso *et al.* 2013) but was historically minimal or absent during reduced
104 sea levels caused by glacial periods of the Pleistocene (Rohling *et al.* 2009), including the most
105 recent glacial maximum (20 to 15 ka; Siddall *et al.* 2003; Ludt & Rocha 2015). Restricted water
106 flow resulted in increased salinity within the Red Sea (Biton *et al.* 2008), leading some to
107 suggest that there was complete extirpation of Red Sea fauna during these periods (Klausewitz
108 1989). The “Pleistocene extirpation” hypothesis, wherein all Red Sea fauna were eliminated
109 during the last glacial maximum (~18 ka) and subsequently re-populated via more recent
110 colonization events, remains controversial and untested with modern comparative approaches
111 (DiBattista *et al.* 2016b), although similar geological events may have occurred in the
112 Mediterranean Sea (Bianchi *et al.* 2012). Thus, despite some agreement on the broad strokes of
113 its geologic history, little consensus has emerged on the processes that shaped the Arabian
114 Peninsula’s present day marine biodiversity, their influence on biodiversity in adjacent regions,
115 and the role of historical closures of the Strait of Bab al Mandab.

116 Butterflyfishes and bannerfishes, brightly colored reef fishes in the family
117 Chaetodontidae, are a potential model system for elucidating the origins, maintenance, and
118 evolutionary history of Red Sea endemics and their influence on species richness in adjacent
119 marine regions. The family is diverse (17 species in the Red Sea and >130 species in the greater
120 Indo-West Pacific; Allen *et al.* 1998) and phylogenetically well-resolved compared to other reef
121 fish families (Cowman 2014). A high proportion of the Chaetodontidae species found in the
122 coastal waters of the Arabian Peninsula are endemic (32%; DiBattista *et al.* 2016a). Although
123 recent molecular phylogenies of chaetodontids have helped to clarify many aspects of their
124 evolutionary history (Fessler & Westneat 2007; Hsu *et al.* 2007; Bellwood *et al.* 2010; Cowman
125 & Bellwood 2011; Cowman & Bellwood 2013; Hodge *et al.* 2014), a lack of sampling of
126 Arabian Peninsula species has impeded our understanding of the diversification in this region.

127 The evolution of endemic species has been linked to ecological traits, such as reductions
128 in dispersal ability and changes in body size (i.e. the island rule; reviewed by Lomolino 2005;
129 Whittaker & Fernandez-Palacios 2007). For reef fishes, certain traits associated with dispersal
130 ability are linked to geographic range size. For example, large, gregarious, and nocturnal species
131 tend to have larger range sizes than small, solitary, and strictly diurnal species (Luiz *et al.* 2012;
132 Luiz *et al.* 2013). Moreover, dispersal ability can potentially influence clade diversification: to
133 successfully colonize and establish populations in peripheral areas, tropical fish species must be
134 good dispersers (Hobbs *et al.* 2012). Following diversification in peripheral areas, newly formed
135 lineages may evolve traits less conducive to dispersal, thus becoming endemic to the area where
136 it originated, as often occurs in the evolution of insular terrestrial endemics (Whittaker &
137 Fernandez-Palacios 2007). We therefore predict that butterflyfishes endemic to the Arabian
138 Peninsula region will have smaller body sizes, higher sociability, and reduced dispersal ability
139 compared to their widespread congeners. Broadly speaking, endemic species tend to be
140 ecological specialists and thus adapted to the environmental condition in which they arose
141 (McKinney 1997). We therefore additionally predict that these endemics will have a higher level
142 of ecological specialization than widespread species. For reef fishes, habitat specialization is
143 often defined by the depth range where they occur and the number of different habitats that they
144 exploit (e.g. coral reefs, rocky reefs, seagrass beds, mangroves; Luiz *et al.* 2012). Dietary
145 specialization is often defined by the proportion of different food categories targeted (Pratchett

146 2014). We predict that butterflyfishes endemic to the Arabian Peninsula region will have higher
147 dietary specialization and reliance on corals for food given recent origins alongside their coral
148 rich habitat (Renema *et al.* 2016). We choose to focus on adult versus larval ecological traits
149 because more information about the former is available, and has been shown to correlate with
150 past (Ottimofiore *et al.* 2017) and present (Luiz *et al.* 2013) geographic range size.

151 The aims of this study are threefold. First, we aim to reconstruct the phylogeny and
152 evolutionary timescale for Red Sea to Arabian Gulf butterflyfishes in order to test whether these
153 peripheral areas intermittently seed the broader Indo-West Pacific with biodiversity
154 (“evolutionary incubator” hypothesis). Outcomes that would allow rejection of this hypothesis
155 include a lack of evidence supporting Arabian Peninsular endemic fish lineages giving rise to
156 Indo-West Pacific fish lineages as well as restricted ancestral ranges expanding into this broader
157 region. Second, we look to test the extent to which butterflyfish maintained a continuous
158 presence in the Red Sea during the major environmental fluctuations of the Pleistocene
159 (“Pleistocene extirpation” hypothesis). Outcomes that would allow rejection of this hypothesis
160 include a lack of evidence supporting Arabian Peninsular endemic fish originating after the
161 glacial cycles of the Pleistocene, as well as colonization events dated only before or after this
162 epoch. Third, we aim to test whether species endemic to the coastal waters of the Arabian
163 Peninsula non-randomly associate with particular ecological traits (“ecological trait” hypothesis),
164 which may be important in explaining patterns of diversification in this region. The expectation
165 here is that endemic fishes are more specialized and thus better adapted to local conditions than
166 their widespread congeners. Outcomes that would allow rejection of this hypothesis include a
167 lack of association between endemism and any of the ecological traits considered here.

168

169 **Materials and Methods**

170 *Materials*

171 Site location, sampling date, and museum voucher information (where available) for each
172 specimen are outlined in Table S1 available as electronic supplementary material. All
173 butterflyfish species included in this study and their geographic distribution are listed in Table 1.
174 As our primary objective is to reconstruct the evolutionary history of butterflyfishes known to
175 occur in the Red Sea and adjacent gulfs or seas, we concentrated our sampling efforts on those

176 species and their closest relatives. Although five major Chaetodontidae lineages were sampled,
177 Chaetodon Clade CH1 (*Chaetodon robustus* and *C. hoefleri*, restricted to the Atlantic; Cowman
178 & Bellwood 2013), and multiple bannerfish genera (*Amphichaetodon*, *Chelmon*, *Chelmonops*,
179 *Coradion*, *Hemitaurichthys*, and *Johnrandallia*) without species represented in the Red Sea were
180 not sampled in this study. Two species of the *Prognathodes* genus were included to facilitate
181 fossil calibration, but were not included in the biogeographic analyses due to their Atlantic
182 distributions (see below).

183 In total, we sampled 47 chaetodontid species (35% of the entire family), which includes
184 all regional endemics and wide-ranging species found in the Arabian Peninsula region, save *Roa*
185 *jayakari*, a rare deepwater species distributed from the Red Sea to coastal India; we were unable
186 to secure a tissue sample as part of this study. Eight of these species have not previously been
187 sampled in phylogenetic studies of the family (Fessler & Westneat 2007; Bellwood 2010;
188 Cowman & Bellwood 2011; Hodge *et al.* 2014). Tissues were preserved in a saturated salt-
189 DMSO solution or 95% ethanol prior to processing. This research was carried out under the
190 general auspices of King Abdullah University of Science and Technology's (KAUST)
191 arrangements for marine research with the Saudi Arabian Coast Guard and the Presidency of
192 Meteorology and Environment. The animal use protocol was approved by KAUST's Biosafety
193 and Ethics Committee (KAUST does not provide specific approval number).

194 195 *Phylogenomics approach*

196 We employ the sequence capture method of ultraconserved elements (UCEs) to produce millions
197 of reads in parallel from multiple butterflyfish specimens collected from the Gulf of Aqaba in the
198 west (Red Sea) to the Hawaiian Archipelago in the east (Pacific Ocean). UCEs are a class of
199 highly conserved and abundant nuclear markers distributed throughout the genomes of most
200 organisms (Bejerano *et al.* 2004; Siepel *et al.* 2005; Reneker *et al.* 2012). These markers do not
201 intersect paralogous genes (Derti *et al.* 2006), do not have retro-element insertions (Simons *et al.*
202 2006), have a range of variant sites (i.e. evolving on different time-scales; Faircloth *et al.* 2012),
203 and have been used to reconstruct phylogenies across vertebrates (Bejerano *et al.* 2004; Faircloth
204 *et al.* 2012, 2013; McCormack *et al.* 2013; Smith *et al.* 2014; Sun *et al.* 2014), including fishes at
205 both shallow (Mcgee *et al.* 2016) and deep (Faircloth *et al.* 2013; Harrington *et al.* 2016; Alfaro

206 *et al.* 2018) phylogenetic scales.

207

208 *DNA library preparation and next-generation sequencing*

209 DNA was extracted with DNeasy Blood and Tissue kits (Qiagen, Valencia, CA), which included
210 an RNase A treatment step. Each extracted sample was visualised by gel electrophoresis to
211 assess DNA quality. Total DNA from each extracted aliquot was quantified using a Qubit
212 dsDNA HS Assay Kit (Invitrogen, Carlsbad, CA), 1.2 µg of DNA per individual sample was
213 fragmented by sonication to 500 base pairs (bp) using a Covaris S2 sonicator (Covaris Inc,
214 Woburn, MA) and used for UCE library prep. In brief, we end-repaired, adenylated, and ligated
215 fragmented DNA to Illumina TruSeq-style adapters, which included custom sequence tags to
216 barcode each individual sample (Faircloth & Glenn, 2012). Following an 18-cycle PCR to
217 amplify indexed libraries for enrichment, we created pools by combining 62.5 ng of eight
218 individual libraries. Each pool was concentrated to 147 ng per µl using a vacuum centrifuge. We
219 then followed an established workflow for target enrichment (Gnirke *et al.* 2009) with
220 modifications specified in Faircloth *et al.* (2012). Specifically, we enriched each pool, targeting
221 UCE loci and their flanking sequence, using synthetic RNA capture probes (MyBaits,
222 Mycroarray, Inc., Ann Arbor, MI). We combined the enriched, indexed pools at equimolar ratios
223 prior to sequencing. The two final pooled libraries were each run paired-end (150 bp sequencing)
224 on independent lanes of an Illumina HiSeq2000 (v3 reagents) at the King Abdullah University of
225 Science and Technology (KAUST) Bioscience Core Lab (BCL). Detailed methods of library
226 enrichment, post-enrichment PCR, and validation using relative qPCR may be found at
227 <http://ultraconserved.org/#protocols>.

228

229 *Sequence read quality control, assembly, and UCE identification*

230 We removed adapter contamination and low quality bases with illumiprocessor (Faircloth, 2013),
231 a parallel wrapper to Trimmomatic (Bolger *et al.* 2014). To assemble the trimmed dataset, we
232 used the PHYLUCE pipeline (version 8ca5884; Faircloth 2016) with the
233 `phyluce_assembly_assemblo_trinity.py` wrapper script for Trinity (version 1.5.0; Grabherr *et al.*
234 2011). We matched assembled contigs to enriched UCE loci by aligning contigs from each
235 species to our UCE probes using the `phyluce_assembly_match_contigs_to_probes.py` script with

236 the LASTZ assembler (Harris 2007). We stored these match results into a SQLite relational
237 database after excluding contigs that matched multiple UCE loci and UCE loci whose probes
238 matched multiple contigs.

239 We used `phyluce_align_seqcap.py` to align UCE loci with MAFFT (Katoh *et al.* 2002;
240 Katoh & Standley 2013). Following alignment, we end- and internally-trimmed alignments with
241 GBLOCKS (Castresana 2000) to improve phylogenetic inference by removing poorly-aligned or
242 highly divergent sites (Talavera & Castresana 2007). We selected loci that were present in at
243 least 75% of our specimens and concatenated the alignments into a PHYLIP-formatted matrix
244 for phylogenetic analysis. We included previously published UCE data for three species in our
245 alignment to represent Acanthomorpha outgroup lineages and more accurately calibrate the
246 phylogeny (see below).

247

248 *Phylogenetic analysis of concatenated UCE data: evaluation of the “evolutionary incubator”*
249 *and “Pleistocene extirpation” hypotheses*

250 We fully partitioned our concatenated alignment by UCE locus and performed Bayesian analyses
251 of the dataset with ExaBayes (Aberer *et al.* 2014) and two independent runs, sampling every 500
252 generations. We used the autostopping convergence criteria of an average standard deviation of
253 split frequencies of less than 5% and visualised the log-likelihood of each chain to ensure
254 convergence in Tracer version 1.6 (Rambaut *et al.* 2014).

255 We estimated divergence times using MCMCTREE in the PAML package on the
256 Bayesian consensus topology. We used the likelihood approximation approach following the
257 two-step procedure described by Dos Reis & Yang (2011) by first estimating a mean substitution
258 rate for the entire alignment with BASEML under a strict molecular clock and then using this
259 estimate to set the `rgene_prior` in MCMCTREE. We used a single, unpartitioned alignment for
260 computational tractability, with an HKY85 model, five categories for the gamma distribution of
261 rate heterogeneity, an `rgene_gamma` prior for the gamma distribution describing gene rate
262 heterogeneity of (2, 371.0575, 1) and a `sigma2_gamma` prior of (2, 5, 1). We adopted a
263 calibration strategy that builds on Harrington *et al.* (2016) by including more proximal
264 acanthomorph outgroups to Chaetodontidae and their immediate relatives. We constrained six
265 nodes on the basis of fossil information using hard lower and soft upper bounds outlined in

266 Figure S1 (available as Supplementary Material). We assigned a minimum amount of prior
267 weight for ages below the lower bound (1e-200) and 5% prior weight for ages higher than the
268 upper bound. Briefly, we link a series of carangimorph, syngnathiform, holocentroid, and
269 lampridiform fossils to the sequences of acanthomorph outgroup fossils as per Harrington et al.
270 (2016). This resulted in the following outgroup node calibrations: acanthuroids versus all other
271 taxa (lower bound: 54.17 Ma; upper bound: 70.84 Ma); acanthurids versus zancids (lower
272 bound: 49.0 Ma; upper bound: 62.7 Ma), *Naso* versus *Acanthurus* (lower bound: 49.0 Ma; upper
273 bound: 57.22 Ma), Chaetodontidae versus Pomacanthidae (lower bound: 29.62 Ma; upper bound:
274 59.26 Ma), and the total-group *Chaetodon* versus *Prognathodes* (lower bound: 7 Ma; upper
275 bound: 47.5 Ma). Further justification for calibrations are available as electronic supplementary
276 material (Appendix S2).

277

278 *Ancestral biogeographic range estimation: evaluation of the “evolutionary incubator” and*
279 *“Pleistocene extirpation” hypotheses*

280 We estimated ancestral distribution patterns for chaetodontid lineages using the pruned time-
281 calibrated phylogeny analysed with the R package BioGeoBEARS (Matzke 2013), which allows
282 several models of biogeographic evolution to be compared via likelihood inference, and the
283 ability to incorporate a parameter allowing for founder-event speciation.. For these analyses, we
284 coded each taxon based on presence/absence in nine discrete geographical areas: Gulf of Aqaba,
285 rest of the Red Sea, Djibouti and Gulf of Aden, Socotra, South Oman, Arabian Gulf, Gulf of
286 Oman and Pakistan, rest of Indian Ocean, and Pacific Ocean. The discrete coding of geographic
287 areas adjacent to the Arabian Peninsula enable a fine-scale investigation of the ancestral
288 biogeography of that region for our taxa of interest. Presence/absence and geographical range
289 data for each taxon were obtained from a combination of DiBattista et al. (2016a) and FishBase
290 (Froese & Pauly 2011). *Prognathodes* spp. (a Chaetodontidae genus) were not considered in this
291 part of the analysis given that these two taxa are restricted to tropical Atlantic waters.

292 We constrained our biogeographic analyses to prohibit colonization events between the
293 Red Sea and Indian/Pacific Ocean regions before 5 Ma reflecting the time when a more
294 permanent connection was formed via the Strait of Bab al Mandab (Bailey 2009). Our
295 BioGeoBEARS analysis evaluated the DEC, DIVALIKE, and BAYAREALIKE models with

296 and without the jump (J) parameter (Matzke 2013). These models describe biogeographic
297 scenarios where dispersal, extinction, cladogenesis, vicariance, and founder events are
298 differentially invoked to explain present day distributional patterns. In our case, we were
299 interested in whether the range-restricted endemics from the coastal waters of the Arabian
300 Peninsula represent ancient relicts, new colonization events, and/or a source of biodiversity (at
301 some point in the past) for the broader Indo-West Pacific.

302

303 *Comparative trait analysis: evaluation of the “ecological trait” hypothesis*

304 In order to determine whether particular species-level traits were associated with the evolution of
305 endemism in this subset of Chaetodontidae species, we fitted a phylogenetic generalized linear
306 model (function ‘*phylglm*’ in R package ‘*phylolm*’ [Ho *et al.* 2016]) that assumed “regional
307 endemism” (i.e. endemic to the coastal waters of the Arabian Peninsula; DiBattista *et al.* 2016a)
308 as the binomial response variable and a suite of ecological traits as the predictive fixed factors.
309 For model selection, we performed a backward stepwise procedure for PGLM’s (function
310 ‘*phylstep*’ in R package ‘*phylolm*’ [Ho *et al.* 2016]), which entailed sequential optimization by
311 removing non-influential fixed-effect terms from the full model based on Akaike information
312 criteria (AIC). Full details on the methods and data sources are provided in Table S2 as
313 electronic supplementary material. We also explore interactions among the predictive traits using
314 a regression tree approach (De’ath & Fabricius 2000; function ‘*rpart*’ in R package ‘*rpart*’
315 [Therneau *et al.* 2015])

316 Among the predictive variables considered were: maximum body size (total length =
317 TL)(Allen *et al.* 1998; Kuitert 2002), depth range inhabited (Allen *et al.* 1998), social structure
318 (three categories ordered from low to high sociability: solitary, pair formation, and group
319 formation; Allen *et al.* 1998; Kuitert 2002; Yabutu & Berumen 2013), habitat breadth (estimated
320 as the sum value of all habitat types inhabited: C = coral, R = rocky, D = deep reef, S = sediment,
321 R = rubble, CO = coastal, CA = algal beds; Allen *et al.* 1998; Kuitert 2002), and dietary reliance
322 on coral reefs (four categories ordered from low to high reliance: planktivore, benthic
323 invertivore, facultative corallivore, and obligate corallivore; Cole & Pratchett 2013). We also
324 included the phylogenetic age of species (Myr) as an additional fixed factor to test if species
325 traits are influenced by time of divergence from sister taxa. For phylogenetic age, we evaluate

326 for each species (regional endemic and widespread) whether we sampled its closest sister species
327 by comparing our phylogeny with those published previously (Cowman & Bellwood 2011) and
328 other published accounts (Kuitert 2002). The ecological traits were selected because they are
329 associated with specialization, fitness, and range expansion in butterflyfishes, and thus may help
330 to explain patterns of evolution in fish endemic to the coral reefs of the Arabian Peninsula. We
331 do note this may be an oversimplification given that our categories are coarse and biased towards
332 adult versus larval traits, which are themselves data deficient. Previous work, however, has
333 demonstrated that traits associated with the successful recruitment of reef fish is more important
334 than traits associated with dispersal in determining differentiation between habitats (Gaither *et*
335 *al.* 2015; Keith *et al.* 2015).

336

337 **Results**

338 *UCE sequences*

339 Reads, contigs, and UCE loci per individual are outlined in Table S3 (available as electronic
340 supplementary material). In summary, we sequenced a total of 153.31 million reads, with a mean
341 of 1.55 million reads per sample from 47 focal taxa (excluding outgroups; also see Table 1).
342 Overall, we assembled a mean of 12,969 contigs (95 CI, min = 10,593, max = 15,345) and 901
343 UCE loci per sample (95 CI, min = 871, max = 932).

344

345 *Phylogenetic reconstruction and timing of divergence to evaluate the “evolutionary incubator”* 346 *and “Pleistocene extirpation” hypotheses*

347 Following assembly, alignment, trimming, and filtering out loci that were present in fewer than
348 75 specimens (for a 75% complete dataset), we retained 971 alignments with a mean length of
349 515.6 bp. The concatenated supermatrix contained 500,642 bp with 52,680 informative sites and
350 was 83.3% complete based on the proportion of non-gap sequences. The following samples were
351 excluded from further analysis due to the low number of loci recovered:

352 *Chaetodon interruptus*1a, *Chaetodon lineolatus*1a, *Chaetodon lunula*1a, and *Chaetodon*
353 *ulietensis*1a (for full details see Table S1); however, tissue replicates were retained for two of the
354 four species listed here (*C. lineolatus* and *C. lunula*).

355 Our Bayesian and maximum likelihood analyses produced a fully resolved topology that

356 shared key points of congruence with prior multi-locus studies of butterflyfishes (Fessler &
357 Westneat 2007; Hsu *et al.* 2007; Bellwood *et al.* 2010; Cowman & Bellwood 2011; Hodge *et al.*
358 2014; see Figure S2). Although direct comparisons to previous phylogenies are difficult because
359 these are missing many of the regional endemics (e.g. *Chaetodon dialeucos*, *C. gardineri*, *C.*
360 *leucopleura*, *C. nigropunctatus*, *C. pictus*, *C. triangulum*, *Heniochus intermedius*), and contain
361 less sequence data and data overlap (e.g. six loci and 73% complete matrix; Hodge *et al.* 2014),
362 where there was overlap in the data sets the tips of the tree displayed similar topologies (Figure
363 S3). In our case, however, almost every node in the tree was strongly supported (posterior
364 probabilities of 1.0; Figure 1 and 2).

365 By only considering a single representative sample per species on our chronogram
366 (Figure 2), we found that the majority of Red Sea to Arabian Gulf butterflyfish diverged from
367 their closest relatives in the last five million years (4.17 Ma to 1.16 Ma), with an average lineage
368 age of 2.39 Ma. In comparison to previous fossil calibrated studies of Chaetodontidae (Cowman
369 & Bellwood 2011; Hodge *et al.* 2014), the mean ages and 95% highest posterior density (HPD)
370 estimates are more restricted, but for the most part overlap with previous estimates (Figure S3).
371 In terms of the topology, although our phylogenetic sampling is restricted, it still captures crown
372 nodes and age estimates of major chaetodontid lineages (with the exception of the bannerfish
373 lineage), as well as subclades containing Red Sea to Arabian Gulf species and their most recent
374 common ancestors (Figure S2). Most of the clades included species pairs diverging with non-
375 overlapping distributions dating back 2 to 1 Ma. This divergence does not appear to coincide
376 with the timing of the emergence of apparent geographic (and geological) barriers such as the
377 Strait of Bab al Mandab (Figure 2 and 3). Regional endemics appear to have diverged earliest
378 from ancestors that gave rise to the clades including *Chaetodon larvatus* and *Chaetodon*
379 *semilarvatus*. At least one entire subclade of CH4 was comprised of regional endemics
380 (*Chaetodon dialeucos*, *C. nigropunctatus*, and *C. mesoleucos*). The split between the
381 butterflyfishes (*Chaetodon*, *Prognathodes*) and bannerfishes (*Heniochus*, *Forcipiger*) was much
382 older, with a mean of 28.7 Ma (95% HPD: 40.0-18.26; Figure 2 and Figure S1), indicating an
383 ancient split between these highly divergent body forms.
384

385 Ancestral range reconstruction to evaluate the “evolutionary incubator” and “Pleistocene
386 extirpation” hypotheses

387 Model comparison revealed the DEC+J model as the most likely (LnL = -250.79, AIC weight =
388 0.76) and the DIVALIKE+J model as the second most likely (LnL = -252.76, AIC weight =
389 0.11; Table 2 and Figure 4). The importance of the J parameter, which models long-distance or
390 “jump” dispersal, is that ancestral ranges often comprise one area rather than several areas. The
391 addition of the J parameter resulted in a significantly better model fit for DEC models when
392 compared via a likelihood ratio test (LRT: $D = 8.67$, $p = 0.0032$).

393 Under the DEC+J model, Chaetodontidae have their crown origins in the Indo-West
394 Pacific, with subsequent dispersal to include the Arabian Peninsula and lineages leading to the
395 base of *Chaetodon* and the bannerfish clade (*Forcipiger/Heniochus*; Figure 4). Within the CH2
396 clade, diversification was restricted to the Pacific Ocean with subsequent dispersal to the Indian
397 Ocean (*Chaetodon madagaskariensis*, *C. punctatofasciaticus*, *C. unimaculatus*), and three of the
398 species have dispersed as far as Socotra (*Chaetodon guttatissimus*, *C. kleinii*, *C. trifasciatus*).
399 Only one species within CH2 diverged in the Gulf of Aden and subsequently colonized the Red
400 Sea (*Chaetodon paucifasciatus*). The age of *C. paucifasciatus* is relatively young (1.5 Ma, HPD:
401 0.8 – 2.3 Ma), suggesting a similar timeline for its occupation of the Red Sea from the Gulf of
402 Aden.

403 In the CH3 clade, three species are present in the Red Sea that were restricted to the
404 Arabian Peninsula (*Chaetodon austriacus*, *C. melapterus*, *C. larvatus*). In the case of sister pair
405 *C. austriacus* and *C. melapterus*, the reconstruction suggests that speciation occurred by
406 vicariance within the Red Sea. Although posterior probabilities make the details of this split
407 equivocal, the most likely scenario is a split between the Gulf of Aqaba and the Red Sea, where
408 *C. austriacus* subsequently recolonized the entire Red Sea but *C. melapterus* expanded out to the
409 Gulf of Aden, Arabian Sea, and Arabian Gulf, but no further. The extended history of the clade,
410 although not completely sampled (see Figure S2), suggests that a widespread ancestor expanded
411 into the Red Sea with subsequent vicariance between the Pacific Ocean, Indian Ocean, and
412 Arabian Peninsula sites. Indeed, *Chaetodon larvatus* appears to originate in Djibouti and the
413 Gulf of Aden followed by dispersal into the Red Sea and South Oman. *Chaetodon trifascialis*, on
414 the other hand, maintained connections across the Indo-West Pacific with subsequent range

415 expansion into the Red Sea.

416 The CH4 clade has been the most successful in terms of butterflyfish colonizing the Red
417 Sea. Eight extant species from CH4 are distributed in at least some parts of the Red Sea
418 (*Chaetodon auriga*, *C. fasciatus*, *C. leucopleura*, *C. lineolatus*, *C. melannotus*, *C. mesoleucos*, *C.*
419 *pictus*, *C. semilarvatus*), four of which are restricted to the Arabian Peninsula (*Chaetodon*
420 *fasciatus*, *C. mesoleucos*, *C. pictus*, *C. semilarvatus*). Moreover, the reconstruction identified a
421 mix of origin states for CH4 species found in the Red Sea. Both *C. fasciatus* and *C. leucopleura*
422 have their origins within the Red Sea, whereas *C. lineolatus* and *C. mesoleucos* have their
423 reconstructed origins at Socotra. The origins of *C. semilarvatus* are placed in South Oman,
424 whereas the origins of *C. pictus* are placed in the Gulf of Oman. With the exception of *C.*
425 *lineolatus*, a widespread Indo-West Pacific species, all CH4 lineages have reconstructed origins
426 in the Arabian Peninsula and Indian Ocean region, and subsequent dispersal was limited from
427 this region. *Chaetodon lineolatus* appears to be the only species in CH4 to originate in the
428 Arabian Peninsula and then disperse across the broader Indo-West Pacific. However, unsampled
429 taxa from this clade are more widely distributed across the Indian and Pacific Oceans (Figure
430 S2).

431 Three taxa of the bannerfish clade are also present in the Red Sea (*Heniochus*
432 *intermedius*, *H. diphreutes*, *Forcipiger flavissimus*), with *H. intermedius* considered a Red Sea to
433 Gulf of Aden endemic. Despite these taxa only being representative of a small proportion of the
434 entire bannerfish clade, the reconstruction suggests a widespread ancestor that diverged in the
435 Arabian Peninsula (*H. intermedius*) with subsequent (successful) colonization of the broader
436 Indo-West Pacific (*F. flavissimus* and *H. diphreutes*).

437
438 *Correlational trait analysis to evaluate the “ecological trait” hypothesis*

439 Based on the best-fit PGLM, depth range and phylogenetic age were negatively correlated with
440 endemism, with depth range being a stronger predictor than phylogenetic age (Table 3, Figure 5
441 and 6). Exploring these relationships using a regression tree approach reveals that the effect of
442 phylogenetic age is dependent on depth range. Endemic species from the Arabian Peninsula
443 region are therefore more likely to be younger than widespread ones, but only for those species
444 with depth ranges extending to mesophotic reefs (depth range > 27 m; Figure 5 and 6).

445 Endemism was not correlated with any of the other factors in the analysis for the butterflyfishes
446 considered here (electronic supplementary material, Table S2 and S4).

447

448 **Discussion**

449 This study used 901 loci to successfully generate a genome-scale phylogeny of bannerfishes and
450 butterflyfishes endemic to the coastal reefs of the Arabian Peninsula. This is the first time this
451 genomic method has been applied to species-level phylogenetic analyses of a reef fish group.

452 Our phylogeny, which includes all shallow water chaetodontid species found in the Red Sea to
453 Arabian Gulf and their close relatives distributed throughout the Indo-West Pacific, provides
454 divergence times with narrow confidence intervals and biogeographic insight into this endemism
455 hotspot. Reconstructing the evolutionary history of these fishes with their widespread relatives
456 does not appear to support the evolutionary incubator hypothesis. That is, despite generating
457 significant biodiversity in the form of endemic species, these peripheral areas of the Arabian
458 Peninsula do not appear to have exported significant biodiversity to the central Indo-West
459 Pacific. In fact, potentially only three species with reconstructed origins in the Arabian Peninsula
460 (*Chaetodon lineolatus*, *Heniochus diphreutes*, *Forcipiger flavissimus*) appear to subsequently
461 disperse to the Indo-West Pacific. Our phylogenetic analyses also revealed that most endemic
462 species originated prior to and persisted through the major environmental fluctuations of the
463 Pleistocene, which does not support the Pleistocene extirpation hypothesis. The ecological trait-
464 based analyses revealed that the evolution of Red Sea to Arabian Gulf endemic butterflyfishes
465 was associated with specialization to shallow reef habitat and, to a lesser extent, species'
466 phylogenetic age.

467

468 *Evaluating the “evolutionary incubator” hypothesis*

469 The Red Sea, Gulf of Aden, Arabian Sea, and Arabian Gulf are all peripheral to the broader
470 Indo-West Pacific biogeographic region and potentially produce/contribute new reef fish species
471 to the centre (see Bowen *et al.* 2013; Hodge *et al.* 2014). Temporally, the Red Sea to Arabian
472 Gulf butterflyfish assemblage (16 species in total) is made up of recently diverged lineages, with
473 ages ranging from 4.17 Ma (*Forcipiger flavissimus*) to 1.16 Ma (*C. austriacus*/*C. melapterus*
474 split). In a few cases, the Red Sea to Gulf of Aden endemics appear to have diverged as the

475 earliest lineage of that clade (e.g. *C. larvatus* and *C. semilarvatus*; Figure 2 and 3). Indeed, the
476 “oldest” endemic butterflyfish lineage in our study, *C. larvatus* (2.86 Ma, 4.3-1.6 Ma 95% HPD),
477 appeared in the late Pliocene, and diverged from an Indo-West Pacific lineage that later gave rise
478 to species allopatric between the two ocean basins (*C. triangulum* in the Indian Ocean and *C.*
479 *baronessa* in the Pacific Ocean). The ancestral range reconstruction of these Arabian Peninsula
480 endemics demonstrates consistent colonization routes to the Red Sea and Arabian Sea via the
481 Indian Ocean from the east (Figure 4), but with few examples of reciprocal expansion from the
482 Arabian Peninsula back to the Indian/Pacific Ocean. For example, both *C. larvatus* and *C.*
483 *semilarvatus* appear to have historically diverged in Djibouti/Gulf of Aden and South Oman,
484 respectively, successfully colonized the Red Sea, but not established further south and east based
485 on present day distributions. Similar reconstruction results were obtained for the regional
486 endemic *C. pictus* (Red Sea to Gulf of Oman), which showed apparent historical divergence in
487 the Gulf of Oman and only recent colonization of the southern limits of the Red Sea.

488 Other endemics appear to have historically diverged within the Red Sea (*C. fasciatus*) or
489 adjacent Djibouti and Gulf of Aden (*C. paucifasciatus*) but not colonized any further to the
490 southeast. Although equivocal based on the probabilistic uncertainty of nodes in the ancestral
491 range reconstruction of the most likely model (DEC+J), there are a number of competing
492 explanations for how *C. austriacus* and *C. melapterus* diverged from each other *within* the
493 coastal waters of the Arabian Peninsula (also see Waldrop *et al.* 2016), particularly since *C.*
494 *melapterus* is the only species in this complex present in the Arabian Gulf. The most likely
495 explanation is based on present day distributions (Figure 3c): *C. austriacus* is largely restricted to
496 the northern and central Red Sea (with rare records in the southern Red Sea and outside of the
497 Red Sea), whereas *C. melapterus* is most abundant within or adjacent to the Arabian Gulf (also
498 with rare records in the southern Red Sea) – these bodies of water show opposite trends in terms
499 of productivity, sea surface temperature, and temporal patterns of environmental variation
500 (Raitsos *et al.* 2013; Pous *et al.* 2015). These environmental conditions are additionally
501 significantly different from the rest of the Indian Ocean, and thus the unique conditions in the
502 Red Sea and Arabian Gulf may help explain how endemics evolved, or at least, concentrated and
503 persisted in these peripheral locations.

504 Despite a lack of supporting evidence for the evolutionary incubator hypothesis, a clear

505 pattern emerges that the unique environmental conditions in these peripheral seas may have
506 contributed to the formation of endemic species as outlined above. For example, some
507 butterflyfish subclades are comprised entirely of regional endemics (e.g. *Chaetodon dialeucos*,
508 *C. mesoleucos*, and *C. nigropunctatus*), which provides further evidence that coral reef habitat
509 surrounding the Arabian Peninsula may have generated a number of new taxa. Moreover, *C.*
510 *dialeucos*, an Omani species, shows geographical divergence with the remaining taxa in its group
511 (Figure 3), which all went on to colonize the Red Sea and the Arabian Gulf and must have
512 therefore encountered contrasting environments at the western and eastern margins of their
513 range. The shallow Arabian Gulf started to fill with seawater approximately 14,000 years ago
514 after being dry during the last glacial maximum prior to that (Lambeck 1996), suggesting that it
515 was seeded by successive waves of colonization from coastal Oman. The same process would
516 have been ongoing at the western margin of the *C. dialeucos* range, except that the conditions
517 encountered in the Red Sea would have contrasted to those in the Arabian Gulf (DiBattista *et al.*
518 2016b). So, while there is some evidence to suggest vicariance at the scale of the Arabian
519 Peninsula (i.e. diversification of most taxa occurred in the Plio-Pleistocene), a stronger scenario
520 is that natural selection driven by the major differences in environment and habitat within the
521 area probably played an important role in the formation of endemic species assemblages (e.g.
522 Gaither *et al.* 2015). Thus, even though the distribution of some of the butterflyfishes considered
523 in the present study does stop abruptly at the entrance of the Strait of Hormuz (*Chaetodon*
524 *collare*, *C. pictus*, and *C. gardneri*), it does not support the argument for geographically-driven
525 allopatry. Indeed, all of these species have a different distributional response near the other end
526 of their distribution at the Strait of Bab al Mandab, which includes stopping before the Straits or
527 extending through the Straits into the southern Red Sea (Figure 3). The alternative is that the
528 incumbent widespread butterflyfish may have restricted the Red Sea to Arabian Gulf endemics
529 from expanding further via competitive exclusion.

530 The current environment of the Red Sea is spatially structured with major contrasts in the
531 cool oligotrophic waters of the northern region compared to the much higher temperatures and
532 productivity of the southern region (i.e. Farasan Islands in Saudi Arabia to the east and Dhalak
533 Archipelago in Eritrea to the west) (Raitsos *et al.* 2013; Racault *et al.* 2015). This shift in
534 environmental conditions is most clearly demonstrated in the differences in life history traits

535 associated with reef fish species that occur in both areas, but is also seen in abundance estimates
536 across these gradients (DiBattista *et al.* 2016a; Roberts *et al.* 2016). Such putative selection
537 gradients are most evident in corals, which show signatures of local adaptation to divergent
538 environmental conditions (D'Angelo *et al.* 2015).

539

540 *Evaluating the “Pleistocene extirpation” hypothesis*

541 The second hypothesis that we tested in this study was the Pleistocene extirpation hypothesis,
542 which predicts that all Red Sea fauna were eliminated during the last glacial maxima (~18 ka)
543 and were only re-populated via recent colonization events (see Hemleben *et al.* 1996; Biton *et al.*
544 2008). The number of species diverging at early stages in the Pleistocene dispute the argument
545 that Red Sea fauna did not survive complete closure or restriction of water flow at the Strait of
546 Bab al Mandab (Figure 2). Although it clearly does not coincide with a single vicariance event
547 given the variability in the splitting dates between closely related species (Figure 3; see
548 Michonneau 2015 for invertebrate examples) and ancestral range reconstruction favoring +J
549 parameter models (i.e. founder events between non-adjacent ocean regions; see Table 2),
550 glaciations likely played a role in their separation. Moreover, even though almost all sister
551 species have small areas of overlap at their range edge, which is usually associated with
552 allopatric speciation, in our case these do not coincide with geographical boundaries (i.e.
553 vicariant chokepoints) such as the Strait of Bab al Mandab (see Figure 3; Lambeck 1996;
554 DiBattista *et al.* 2016b). In fact, the non-congruent age and distribution of the endemic species
555 indicate a series of variable events, which may reflect localized patterns of habitat and
556 environmental change as outlined in the previous section. The best example is the relatively
557 young clade of Arabian Peninsula endemics: *C. dialeucos*, *C. nigropunctatus*, and *C. mesoleucos*
558 (crown node age 2.0 Ma; 2.9-1.2 Ma 95% HPD). This group appears to have been influenced by
559 boundaries presented by the Omani coastline across areas where there are known changes in the
560 upwelling regime (Shi *et al.* 2000; McIlwain *et al.* 2005). This is in sharp contrast to the Indo-
561 West Pacific parrotfishes, where present day species boundaries support the notion of allopatric
562 divergence (Choat *et al.* 2012), and endemics appear to have diverged into one or more
563 subsequent endemics (i.e. secondary endemism; Rotondo *et al.* 1981) based on sympatrically
564 distributed sister-species pairs (highlighted in Choat *et al.* 2012). Moreover, Red Sea endemics

565 from most other families of reef fish appear to have equal proportions of allopatrically and
566 sympatrically distributed sister-species (Hodge *et al.* 2014), which is not the case for the
567 butterflyfishes.

568 The diversification of these butterflyfishes occurred at a time when the coral assemblages
569 of the world's reefs underwent a major change in coral composition and growth forms. The
570 global proportion of staghorn coral occurrences in coral assemblages persisted throughout most
571 of the Cenozoic but increased substantially during the Pliocene and especially the Quaternary
572 (Renema *et al.* 2016). Indeed, the rapidly growing branching acroporid corals offered different
573 structural components in terms of shelter and feeding/foraging modes when compared to massive
574 corals such as poritids that dominated Miocene reefs more than 5 Ma. Thus, the chaetodontids of
575 the Arabian Peninsula (particularly the corallivorous species) were exposed to a much more
576 dynamic environment than the widespread Indo-West Pacific species (Coles 2003) because of
577 their close association with sensitive coral genera that proliferated in the region.

578

579 *Evaluating the “ecological trait” hypothesis*

580 The third hypothesis that we test here is whether ecological traits are linked to the evolution of
581 endemism among butterflyfishes in the Red Sea to Arabian Gulf. We found a negative,
582 significant relationship between endemism and depth range and, to a lesser extent, phylogenetic
583 age for these butterflyfishes (Figure 5 and 6). The relationship between a narrow versus broad
584 depth range and endemism supports the view that endemic species tend to be more specialized to
585 local resources than widespread species (Hawkins *et al.* 2000). The majority of regional
586 endemics in this study had depth ranges that did not extend deeper than 25 meters (Figure 6),
587 despite the availability of light dependent coral habitat extending beyond that (Kahng *et al.*
588 2010). The broad range of ages represented by these shallow water specialists suggests that
589 adaptation to shallow reefs occurred multiple times across a relatively wide time frame (i.e. 1.3
590 to 3.3 Myr). On the other hand, speciation of endemics with a preference for deep reefs seems to
591 be a recent phenomenon, as deeper depth ranges were strongly associated with young age (less
592 than 1.75 Myr; Figure 6).

593

594 *Comments on incomplete sampling and biogeographic biases*

595 The goal of this study was to reconstruct the evolutionary history of Red Sea to Arabian Gulf
596 butterflyfishes. As is the case with all phylogenetic and biogeographic reconstructions, our
597 results have to be interpreted in light of the taxa that are not sampled, both extant and extinct.
598 Indeed, the inclusion of missing taxa has the potential to alter lineage relationships and their age
599 estimates, whereas their geographic distribution may alter the most likely biogeographic
600 scenarios reconstructed across the tree (see discussion in Cowman & Bellwood 2013). Here, we
601 have been able to sample all Red Sea to Arabian Gulf butterflyfishes (save one species, *Roa*
602 *jayakari*), and their close relatives from the Indian and Pacific Oceans, across four major
603 chaetodontid lineages (Figure S2). From a temporal perspective, the topology and ages estimated
604 for the genomic scale UCE data overlap with previous studies (Figures S2 and S3). Moreover,
605 our sampling of eight species that have not previously been included in phylogenetic studies of
606 the Chaetodontidae family means that for 13 out of the 17 Arabian Peninsula species, we are
607 confident that we have sampled their direct sister lineage. Two of the outstanding three species
608 (*Chaetodon melannotus*, *C. trifascialis*) are wide ranging Indo-West Pacific taxa that are
609 reconstructed to have dispersed to the Arabian Peninsula (Figure 4). The most likely sister
610 species of *C. melannotus* is *C. ocellicaudus* (Kuitert 2002; also see Figure S2), a west Pacific
611 species not sampled in our dataset. In the case of *C. trifascialis*, it is placed as the sister lineage
612 of a subclade of CH3 containing 10 species distributed across the Indian and Pacific Oceans, of
613 which we sampled four (Figure S2; Cowman & Bellwood 2011). The final outstanding species,
614 *Chaetodon leucopleura*, is placed as a sister species to *Chaetodon gardineri*. Both species have
615 not previously been sampled in phylogenetic studies, but are recognized to be closely related to a
616 third species, *Chaetodon selene* (widespread in the west Pacific, Kuitert 2002), which was not
617 sampled in our UCE dataset. In each of these three cases, and more broadly across the family, the
618 inclusion of unsampled species would increase the influence of the Indian Ocean and Pacific
619 Ocean in the ancestral estimation of biogeographic ranges. As such, it would act to strengthen
620 our conclusion that even though the Red Sea and adjacent gulfs and seas have been important for
621 the generation of endemic species, they have had little contribution to the wider Indo-West
622 Pacific diversity of butterflyfishes.

623

624 *Conclusion*

625 It appears that the unique environmental conditions in the coastal waters of the Arabian
626 Peninsula may have contributed to the formation of endemic butterflyfishes; however, there is a
627 lack of evidence for endemics contributing significant species richness to adjacent seas (i.e.
628 evolutionary incubator hypothesis). Moreover, even with catastrophic environmental instability
629 experienced by the Red Sea and coastal environments of the Arabian Peninsula due to sea level
630 changes associated with glacial cycles (Ludt & Rocha 2015), there is no evidence for a massive
631 extirpation of butterflyfish fauna in the region (i.e. Pleistocene extirpation hypothesis; also see
632 DiBattista *et al.* 2016b). The broad range of phylogenetic ages among endemic, shallow water
633 butterflyfishes supports the view that species may have survived in isolated refugia within the
634 Red Sea (DiBattista *et al.* 2016b). None of the dispersal-related traits were associated with
635 endemism, suggesting that factors other than those related to species intrinsic dispersal potential
636 may be limiting dispersal into the greater Indian Ocean (e.g. coastline geography, oceanographic
637 barriers).

638

639 **Acknowledgements**

640 This work was supported by the KAUST Office of Competitive Research Funds (OCRF) under
641 Award No. CRG-1-2012-BER-002 and baseline research funds to M.L.B.; National Geographic
642 Society Grant 9024-11 to J.D.D.; National Science Foundation grant DEB-0842397 to M.E.A.;
643 California Academy of Sciences funding to L.A.R; Australia Research Council Discovery ECR
644 Award DE170100516 to P.F.C. For support in Socotra, we kindly thank the Ministry of Water
645 and Environment of Yemen, staff at the Environment Protection Authority (EPA) Socotra, and
646 especially Salah Saeed Ahmed, Fouad Naseeb and Thabet Abdullah Khamis, as well as Ahmed
647 Issa Ali Affrar from Socotra Specialist Tour for handling general logistics. For logistic support
648 elsewhere, we thank Eric Mason at Dream Divers, Nicolas Prévot at Dolphin Divers and the
649 crew of the M/V Deli in Djibouti, David Pence, the KAUST Coastal and Marine Resources Core
650 Lab and Amr Gusti, the Administration of the British Indian Ocean Territory and Chagos
651 Conservation Trust, the Ministry of Agriculture and Fisheries in Oman including Abdul Karim,
652 the Ministère de la Pêche et des Ressources Halieutiques – Madagascar, the Western Australia
653 Department of Fisheries, Parks Australia, as well as the University of Milano-Bicocca Marine
654 Research and High Education Centre in Magoodhoo, the Ministry of Fisheries and Agriculture,

655 Republic of Maldives, and the community of Maghoodhoo, Faafu Atoll. For specimen
656 collections, we thank David Bellwood, Brian Bowen, John Burns, Darren Coker, Richard
657 Coleman, Joshua Copus, Joshua Drew, Iria Fernandez-Silva, Michelle Gaither, Brian Greene,
658 Elliott Jessup, Randy Kosaki, Jason Leonard, Keo Lopes, Sarah Longo, Cassie Lyons, Jennifer
659 McIlwain, Gerrit Nanninga, David Pence, Mark Priest, Richard Pyle, Frédéric Ramahatratra,
660 Mark Royer, Pablo Saenz-Agudelo, Anne Sheppard, Charles Sheppard, Jacqueline Troller,
661 Daniel Wagner, Robert Whitton, and members of the Reef Ecology Lab at KAUST. For
662 assistance with bench work at KAUST we thank Craig Michell. We also acknowledge important
663 contributions from Robert J. Toonen, John E. Randall, Jo-Ann C. Leong, David Catania for
664 assistance with specimen archiving, and the KAUST Bioscience Core Laboratory with
665 Sivakumar Neelamegam and Hicham Mansour for their assistance with Illumina sequencing.
666 Special thanks to Scott Partridge for the use of his illustrations of the *Chaetodon* genus.

667

668 **References**

669 Aberer AJ, Kobert K, Stamatakis A (2014) ExaBayes: Massively parallel Bayesian tree inference
670 for the whole-genome era. *Molecular Biology and Evolution* **31**, 2553–2556.

671

672 Allen GR, Steene R, Allen M. (1998) A Guide to Angelfishes and Butterflyfishes. 250 p.
673 Odyssey Publishing, Sydney.

674

675 Alfaro ME, Faircloth BC, Harrington RC, Sorenson L, Friedman M, Thacker CE, Oliveros CH,
676 Černý D, Near TJ (2018) Explosive diversification of marine fishes at the Cretaceous–
677 Palaeogene boundary. *Nature Ecology & Evolution* DOI: 10.1038/s41559-018-0494-6.

678

679 Bailey G (2009) The Red Sea, Coastal Landscapes, and Hominin Dispersals in *The Evolution of*
680 *Human Populations in Arabia Paleoenvironments, Prehistory and Genetics*. Ed Petraglia MD,
681 Springer Science.

682

683 Bejerano G, Haussler D, Blanchette M (2004) Into the heart of darkness: large-scale clustering
684 of human non-coding DNA. *Bioinformatics* **20**, i40-i48.

685
686 Bellwood DR, Klanten S, Cowman PF, Pratchett MS, Konow N, van Herwerden L (2010)
687 Evolutionary history of the butterflyfishes (f: Chaetodontidae) and the rise of coral feeding
688 fishes. *Journal of Evolutionary Biology* **23**, 335-349.
689
690 Bianchi CN, Morri C, Chiantore M, Montefalcone M, Parravicini V, Rovere A (2012)
691 Mediterranean Sea biodiversity between the legacy from the past and a future of change. In *Life*
692 *in the Mediterranean Sea: a look at habitat changes*, 1-55.
693
694 Biton E, Gildor H, Peltier WR (2008) Red Sea during the Last Glacial Maximum: Implications
695 for sea level reconstruction. *Paleoceanography* **23**.
696
697 Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: A flexible trimmer for Illumina Sequence
698 Data. *Bioinformatics* **30**, 2114-2120.
699
700 Bowen BW, Rocha LA, Toonen RJ, Karl SA (2013) The origins of tropical marine biodiversity.
701 *Trends in Ecology and Evolution* **28**, 359-366.
702
703 Castresana J (2000) Selection of conserved blocks from multiple alignments for their use in
704 phylogenetic analysis. *Molecular Biology and Evolution* **17**, 540-552.
705
706 Choat JH, Klanten OS, van Herwerden L, Robertson DR, Clements KD (2012) Patterns and
707 processes in the evolutionary history of parrotfishes (Family Labridae). *Biological Journal of the*
708 *Linnean Society* **107**, 529-557.
709
710 Cole AJ, Pratchett MS (2014) Diversity in diet and feeding behaviour of butterflyfishes: reliance
711 on reef corals versus reef habitats. In: Pratchett MS, Berumen ML, Kapoor BG (eds) *Biology of*
712 *Butterflyfishes*. CRC Press, Boca Raton, FL, USA, pp. 107-139.
713
714 Cole AJ, Pratchett MS, Jones GP (2008) Diversity and functional importance of coral-feeding

715 fishes on tropical coral reefs. *Fish and Fisheries* **9**, 286-307.
716
717 Coles SL (2003) Coral species diversity and environmental factors in the Arabian Gulf and the
718 Gulf of Oman: a comparison to the Indo-Pacific region. *Atoll Research Bulletin* **507**, 1-19.
719
720 Cowman PF (2014) Historical factors that have shaped the evolution of tropical reef fishes: A
721 review of phylogenies, biogeography, and remaining questions. *Frontiers in Genetics* **5**, 1–15.
722
723 Cowman PF, Bellwood DR (2011) Coral reefs as drivers of cladogenesis: expanding coral reefs,
724 cryptic extinction events, and the development of biodiversity hotspots. *Journal of Evolutionary*
725 *Biology* **24**, 2543-2562.
726
727 Cowman PF, Bellwood DR (2013) The historical biogeography of coral reef fishes: Global
728 patterns of origination and dispersal. *Journal of Biogeography* **40**, 209-224.
729
730 Cowman PF, Parravicini V, Kulbicki M, Floeter SR (2017) The biogeography of tropical reef
731 fishes: endemism and provinciality through time. *Biological Reviews* **92**, 2112-2130.
732
733 Crosby MP, Reese ES, Berumen ML (2013) Butterflyfishes as ambassadors of coral reefs.
734 Chapter 10 In Pratchett, Berumen, Kapoor *Biology of Butterflyfishes*.
735
736 D'angelo C, Hume BC, Burt J, Smith EG, Achterberg EP, Wiedenmann J (2015) Local
737 adaptation constrains the distribution potential of heat-tolerant *Symbiodinium* from the
738 Persian/Arabian Gulf. *The ISME Journal* **9**, 2551-2560.
739
740 De'ath G, Fabricius KE (2000) Classification and regression trees: a powerful yet simple
741 technique for ecological data analysis. *Ecology* **81**, 3178-3192.
742
743 Derti A, Roth FP, Church GM, Wu CT (2006) Mammalian ultraconserved elements are
744 strongly depleted among segmental duplications and copy number variants. *Nature*

745 *Genetics* **38**, 1216-1220.

746

747 DiBattista JD, Wilcox C, Craig MT, Rocha LA, Bowen BW (2010) Phylogeography of the
748 Pacific Blueline Surgeonfish, *Acanthurus nigroris*, reveals high genetic connectivity and a
749 cryptic endemic species in the Hawaiian Archipelago. *Journal of Marine Biology* 2011.

750

751 DiBattista JD, Berumen ML, Gaither MR, Rocha LA, Eble JA, Choat JH, Craig MT,
752 Skillings DJ, Bowen BW (2013) After continents divide: comparative phylogeography of reef
753 fishes from the Red Sea and Indian Ocean. *Journal of Biogeography* **40**, 1170-1181.

754

755 DiBattista JD, Roberts M, Bouwmeester J, Bowen BW, Coker DF, Lozano-Cortés DF,
756 Choat JH, Gaither MR, Hobbs JP, Khalil M, et al. (2016a) A review of contemporary patterns of
757 endemism for shallow water reef fauna in the Red Sea. *Journal of Biogeography* **43**, 423-439.

758

759 DiBattista JD, Choat JH, Gaither MR, Hobbs JP, Lozano-Cortés DF, Myers RF,
760 Paulay G, Rocha LA, Toonen RJ, Westneat M, et al. (2016b) On the origin of endemic species in
761 the Red Sea. *Journal of Biogeography* **43**, 13-30.

762

763 Dos Reis M, Yang Z (2011) Approximate likelihood calculation on a phylogeny for
764 Bayesian estimation of divergence times. *Molecular Biology and Evolution* **28**, 2161-2172.

765

766 Eble JA, Toonen RJ, Sorenson L, Basch LV, Papastamatiou YP, Bowen BW (2011) Escaping
767 paradise: larval export from Hawaii in an Indo-Pacific reef fish, the yellow tang (*Zebrasoma*
768 *flavescens*). *Marine Ecology Progress Series* **428**, 245.

769

770 Faircloth BC (2013) illumiprocessor: a trimmomatic wrapper for parallel adapter and quality
771 trimming. DOI: 10.6079/J9ILL.

772

773 Faircloth BC. (2016) PHYLUCE is a software package for the analysis of conserved genomic
774 loci. *Bioinformatics* **32**, 786-788.

775
776 Faircloth BC, Glenn TC (2012) Not all sequence tags are created equal: designing and validating
777 sequence identification tags robust to indels. *PLOS ONE* **7**, e42543.
778
779 Faircloth BC, McCormack JE, Crawford NG, Harvey MG, Brumfield RT, Glenn
780 TC (2012) Ultraconserved elements anchor thousands of genetic markers spanning multiple
781 evolutionary timescales. *Systematic Biology* **61**, 1004.
782
783 Faircloth BC, Sorenson L, Santini F, Alfaro ME (2013) A phylogenomic perspective on the
784 radiation of ray-finned fishes based upon targeted sequencing of ultraconserved elements
785 (UCEs). *PLOS ONE* **8**, e65923.
786
787 Fessler JL, Westneat MW (2007) Molecular phylogenetics of the butterflyfishes
788 (Chaetodontidae): taxonomy and biogeography of a global coral reef fish family. *Molecular*
789 *Phylogenetics and Evolution* **45**, 50-68.
790
791 Froese R, Pauly D (2011) FishBase: World Wide Web electronic publication. Available at:
792 www.fishbase.org (accessed February, 2016).
793
794 Gaither MR, Toonen RJ, Robertson DR, Planes S, Bowen BW (2010) Genetic evaluation of
795 marine biogeographical barriers: perspectives from two widespread Indo-Pacific snappers
796 (*Lutjanus kasmira* and *Lutjanus fulvus*). *Journal of Biogeography* **37**, 133-147.
797
798 Gaither MR, Jones SA, Kelley C, Newman SJ, Sorenson L, Bowen BW (2011)
799 High connectivity in the deepwater snapper *Pristipomoides filamentosus* (Lutjanidae) across the
800 Indo-Pacific with isolation of the Hawaiian Archipelago. *PLOS ONE* **6**, e28913.
801
802 Gaither MR, Bernal MA, Coleman RR, Bowen BW, Jones SA, Simison WB, Rocha LA (2015)
803 Genomic signatures of geographic isolation and natural selection in coral reef fishes. *Molecular*
804 *Ecology* **24**, 1543-1557.

805
806 Girdler RW, Styles P (1974) Two-stage Red Sea floor spreading. *Nature* **247**, 7-11.
807
808 Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan
809 L, Raychowdhury R, Zeng Q, et al. (2011) Full-length transcriptome assembly from RNA-seq
810 data without a reference genome. *Nature Biotechnology* **29**, 644-652.
811
812 Harrington RC, Faircloth BC, Eytan RI, Smith WL, Near TJ, Alfaro ME, Friedman M (2016)
813 Phylogenomic analysis of carangimorph fishes reveals flatfish asymmetry arose in a blink of the
814 evolutionary eye. *BMC Evolutionary Biology* **16**, 224.
815
816 Harris RS (2007) Improved pairwise alignment of genomic DNA. Ph.D. Thesis. Pennsylvania
817 State University.
818
819 Hawkins JP, Roberts CM, Clark V (2000) The threatened status of restricted-range coral reef fish
820 species. *Animal Conservation* **3**, 81-88.
821
822 Ho LST, Ane C, Lachlan R, Tarpinian K, Feldman R, Ho MLST (2016) Package ‘phylolm’.
823
824 Hobbs JPA, Jones GP, Munday PL, Connolly SR, Srinivasan M (2012) Biogeography and
825 the structure of coral reef fish communities on isolated islands. *Journal of Biogeography* **39**,
826 130-139.
827
828 Hodge JR, Herwerden L, Bellwood DR (2014) Temporal evolution of coral reef fishes: global
829 patterns and disparity in isolated locations. *Journal of Biogeography* **41**, 2115-2127.
830
831 Hsu KC, Chen JP, Shao KT (2007) Molecular phylogeny of *Chaetodon* (Teleostei:
832 Chaetodontidae) in the Indo-West Pacific: evolution in geminate species pairs and species
833 groups. *Raffles Bulletin of Zoology* **14**, 77-86.
834

835 Hubert-Ferrari A, King G, Manighetti I, Armijo R, Meyer B, Tapponnier P (2003)
836 Long-term elasticity in the continental lithosphere; modelling the Aden Ridge propagation and
837 the Anatolian extrusion process. *Geophysical Journal International* **153**, 111-132.
838
839 Kahng SE, Garcia-Sais JR, Spalding HL, Brokovich E, Wagner D, Weil E, Hinderstein L,
840 Toonen RJ (2010) Community ecology of mesophotic coral reef ecosystems. *Coral Reefs* **29**,
841 255-275.
842
843 Katoh K, Misawa K, Kuma K, Miyata T (2002) MAFFT: a novel method for rapid multiple
844 sequence alignment based on fast Fourier transform. *Nucleic Acids Research* **30**, 3059-3066.
845
846 Katoh K, Standley DM. 2013 MAFFT multiple sequence alignment software version 7:
847 improvements in performance and usability. *Molecular Biology and Evolution* **30**, 772-780.
848
849 Kemp J (1998) Zoogeography of the coral reef fishes of the Socotra Archipelago. *Journal of*
850 *Biogeography* **25**, 919-933.
851
852 Keith SA, Woolsey ES, Madin JS, Byrne M, Baird AH (2015) Differential establishment
853 potential of species predicts a shift in coral assemblage structure across a biogeographic barrier.
854 *Ecography* **38**, 1225-1234.
855
856 Klausewitz W (1989) Evolutionary history and zoogeography of the Red Sea ichthyofauna.
857 *Fauna of Saudi Arabia* **10**, 310-337.
858
859 Kuitert RH (2002) Butterflyfishes, Bannerfishes and their Relatives: A Comprehensive Guide to
860 Chaetodontidae and Microcanthidae. TMC Publishing, Chorleywood, UK.
861
862 Lambeck K (1996) Shoreline reconstructions for the Persian Gulf since the last glacial
863 maximum. *Earth and Planetary Science Letters* **142**, 43-57.
864

865 Lomolino MV (2005) Body size evolution in insular vertebrates: generality of the island rule.
866 *Journal of Biogeography* **32**, 1683-1699.
867

868 Ludt WB, Rocha LA (2015) Shifting seas: the impacts of Pleistocene sea-level fluctuations on
869 the evolution of tropical marine taxa. *Journal of Biogeography* **42**, 25-38.
870

871 Luiz OJ, Madin JS, Robertson DR, Rocha LA, Wirtz P, Floeter SR (2012) Ecological traits
872 influencing range expansion across large oceanic dispersal barriers: insights from tropical
873 Atlantic reef fishes. *Proceedings of the Royal Society B* **279**, 1033-1040.
874

875 Luiz OJ, Allen AP, Robertson DR, Floeter SR, Kulbicki M, Vigliola L, Becheler R,
876 Madin JS (2013) Adult and larval traits as determinants of geographic range size among tropical
877 reef fishes. *Proceedings of National Academy of Sciences U.S.A.* **110**, 16498-16502.
878

879 Malay MCMD, Paulay G (2010) Peripatric speciation drives diversification and distributional
880 pattern of reef hermit crabs (Decapoda: Diogenidae: *Calcinus*). *Evolution* **64**, 634-662.
881

882 Matzke NJ (2013) Probabilistic historical biogeography: new models for founder-event
883 speciation, imperfect detection, and fossils allow improved accuracy and model-testing.
884 *Frontiers in Biogeography* **4**, 210.
885

886 McCormack JE, Harvey MG, Faircloth BC, Crawford NG, Glenn TC, Brumfield RT (2013) A
887 phylogeny of birds based on over 1,500 loci collected by target enrichment and high-throughput
888 sequencing. *PLOS ONE* **8**, e54848.
889

890 McGee MD, Faircloth BC, Borstein SR, Zheng J, Darrin Hulsey C, Wainwright PC, Alfaro ME
891 (2016) Replicated divergence in cichlid radiations mirrors a major vertebrate innovation.
892 *Proceedings of the Royal Society B* **283**.
893

894 McIlwain JL, Claereboudt MR, Al-Oufi HS, Zaki S, Goddard JS (2005) Spatial variation in age

895 and growth of the kingfish (*Scomberomorus commerson*) in the coastal waters of the Sultanate of
896 Oman. *Fisheries Research* **73**, 283-298.

897

898 McKinney ML (1997) Extinction vulnerability and selectivity: combining ecological and
899 paleontological views. *Annual Review of Ecology and Systematics* **28**, 495-516.

900

901 Michonneau F (2015) Cryptic and not-so-cryptic species in the complex “*Holothuria*
902 (*Thymiosycia*) *imaptiens*” (Forsskål, 1775)(Echinodermata: Holothuroidea: Holothuriidae).
903 *bioRxiv*. DOI: 10.1101/014225.

904

905 Ottimofiore, E, Albouy, C, Leprieur, F, Descombes, P, Kulbicki, M, Mouillot, D, Parravicini, V,
906 Pellissier L (2017) Responses of coral reef fishes to past climate changes are related to life-
907 history traits. *Ecology and Evolution* **7**, 1996-2005.

908

909 Pous S, Lazure P, Carton X (2015) A model of the general circulation in the Persian Gulf and in
910 the Strait of Hormuz: Intraseasonal to interannual variability. *Continental Shelf Research* **94**, 55-
911 70.

912

913 Pratchett MS (2014) Feeding preferences and dietary specialisation among obligate coral-
914 feeding butterflyfishes. In: Pratchett MS, Berumen ML, Kapoor BG (eds) *Biology of*
915 *Butterflyfishes*. CRC Press, Boca Raton, FL, USA, pp. 140–179.

916

917 Racault MF, Raitsos DE, Berumen ML, Brewin RJ, Platt T, Sathyendranath S, Hoteit I (2015)
918 Phytoplankton phenology indices in coral reef ecosystems: application to ocean-color
919 observations in the Red Sea. *Remote Sensing of Environment* **160**, 222-234.

920

921 Raitsos DE, Pradhan Y, Brewin RJW, Stenchikov G, Hoteit I (2013) Remote sensing the
922 phytoplankton seasonal succession of the Red Sea. *PLOS ONE* **8**, e64909.

923

924 Rambaut A, Suchard MA, Xie D, Drummond AJ (2014) Tracer v1.6,

925 <http://beast.bio.ed.ac.uk/Tracer>.

926

927 Reneker J, Lyons E, Conant GC, Pires JC, Freeling M, Shyu CR, Korkin D (2012) Long
928 identical multispecies elements in plant and animal genomes. *Proceedings of the National
929 Academy of Sciences U.S.A.* **109**, E1183-E1191.

930

931 Roberts MB, Jones GP, McCormick MI, Munday PL, Neale S, Thorrold S, Robitzsch VS,
932 Berumen ML (2016) Homogeneity of coral reef communities across 8 degrees of latitude in the
933 Saudi Arabian Red Sea. *Marine Pollution Bulletin* **105**, 558-565.

934

935 Rohling EJ, Grant K, Bolshaw M, Roberts AP, Siddall M, Hemleben C, Kucera M (2009)
936 Antarctic temperature and global sea level closely coupled over the past five glacial cycles.
937 *Nature Geoscience* **2**, 500-504.

938

939 Schuster F (2002) Oligocene and Miocene examples of *Acropora*-dominated
940 palaeoenvironments: Mesohellenic Basin (NW Greece) and northern Gulf of Suez (Egypt).
941 *Proceedings of the 9th International Coral Reef Symposium Bali* **1**, 199-204.

942

943 Shi W, Morrison JM, Bohm E, Manghnani V (2000) The Oman upwelling zone during 1993,
944 1994 and 1995. *Deep-Sea Research II* **47**, 1227-1247.

945

946 Siddall M, Rohling EJ, Almogi-Labin A, Hemleben C, Meischner D, Schmelzer I, Smeed
947 DA (2003) Sea-level fluctuations during the last glacial cycle. *Nature* **423**, 853-858.

948

949 Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K, Clawson H,
950 Spieth J, Hillier LW, Richards S, et al. (2005) Evolutionarily conserved elements in vertebrate,
951 insect, worm, and yeast genomes. *Genome Research* **15**, 1034-1050.

952

953 Simons C, Pheasant M, Makunin IV, Mattick JS (2006) Transposon-free regions in
954 mammalian genomes. *Genome Research* **16**, 164-172.

955
956 Skillings DJ, Bird CE, Toonen RJ (2010) Gateways to Hawai‘i: genetic population
957 structure of the tropical sea cucumber *Holothuria atra*. *Journal of Marine Biology* 2011.
958
959 Smeed D (1997) Seasonal variation of the flow in the strait of Bah al Mandab. *Acta*
960 *Oceanologica* **20**, 773-781.
961
962 Smith BT, McCormack JE, Cuervo AM, Hickerson MJ, Aleixo A, Cadena CD, Pérez-
963 Emán J, Burney CW, Xie X, Harvey MG, et al. (2014) The drivers of tropical speciation. *Nature*
964 **515**, 406-409.
965
966 Sun K, Meiklejohn KA, Faircloth BC, Glenn TC, Braun EL, Kimball RT (2014) The evolution
967 of peafowl and other taxa with ocelli (eyespot): a phylogenomic approach. *Proceedings of the*
968 *Royal Society B* **281**, 20140823.
969
970 Talavera G, Castresana J (2007) Improvement of phylogenies after removing divergent and
971 ambiguously aligned blocks from protein sequence alignments. *Systematic Biology* **56**, 564-577.
972
973 Therneau T, Atkinson B, Ripley B (2015) ‘rpart’. R package version 4.1-10.
974
975 Whittaker RJ, Fernández-Palacios JM (2007) Island biogeography: ecology, evolution, and
976 conservation. Oxford University Press.
977
978 Yabuta S, Berumen ML (2013) Social structures and spawning behaviour of *Chaetodon*
979 butterflyfishes. In *Biology of Butterflyfishes*, 200-225.

980
981
982
983

984 **Figure Legends**

985 **Figure 1.** Inferred phylogeny of Red Sea to Arabian Gulf butterflyfish species, including some
986 of closest their congeners, based on ExaBayes analysis of ultraconserved element (UCE) data.
987 Yellow dots on node labels indicate a posterior probability of 1, whereas grey dots indicate a
988 posterior probability of less than 1 but greater than 0.6. Clades based on Bellwood et al. (2010)
989 and Cowman & Bellwood (2011) are indicated. Records for each species are mapped onto the
990 topology as follows: red = Red Sea to Arabian Gulf, green = rest of Indian Ocean, and blue =
991 Pacific Ocean (PO).

992
993 **Figure 2.** A fossil calibrated chronogram for select Chaetodontidae species based on analysis of
994 ultraconserved element (UCE) data. The time scale is calibrated in millions of years before
995 present. Node ages are presented as median node heights with 95% HPD intervals represented by
996 bars. Significant geological events in the coastal waters of the Arabian Peninsula are temporally
997 indicated by red dashed lines.

998
999 **Figure 3.** Distributions, range overlap, and ages of divergence in eight clades of butterflyfish
1000 from the *Chaetodon* genus that contain species inhabiting the Red Sea to Arabian Gulf region.
1001 Clade structure and node ages (median node heights with 95% HPD intervals represented by
1002 bars) were extracted from Figure 2.

1003
1004 **Figure 4.** Ancestral range estimations inferred using the DEC + J model based on a time
1005 calibrated Bayesian phylogeny of Chaetodontidae species. States at branch tips indicate the
1006 current geographical distributions of taxa, whereas states at nodes indicate the inferred ancestral
1007 distributions before speciation (middle) and after (corner). The regions considered in this
1008 analysis include: Gulf of Aqaba, rest of Red Sea, Djibouti and Gulf of Aden, Socotra, South
1009 Oman, Arabian Gulf, Gulf of Oman and Pakistan, rest of Indian Ocean, and Pacific Ocean.
1010 Abbreviations: Plio. = Pliocene; Ple. = Pleistocene. Significant vicariance in the Red Sea to
1011 Arabian Gulf region are indicated by red dashed lines.

1012
1013 **Figure 5.** Estimated probability of endemism among Red Sea to Arabian Gulf butterflyfish
1014 species, including some of their closest congeners, as a function of depth range. Different line

1015 types represent variability in estimated species phylogenetic age extracted from Figure 2 (see
1016 legend).

1017

1018 **Figure 6.** The classification of species-level traits associated with endemism among the Red Sea
1019 to Arabian Gulf butterflyfishes (a). Data on the leaves (represented by squares) provide the
1020 probability of endemism (top) and the percentage of all observations in the node (bottom). The
1021 right panel shows the prediction surface (b).

1022

1023 **Data accessibility**

1024 Data associated with this manuscript are available under NCBI BioProject PRJNA484421,
1025 available at <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA484421>.

1026

1027 **Authors' contributions**

1028 J.D.D., M.E.A., L.A.R., J.H.C., and M.L.B. designed the study; J.D.D., L.S., J.P.A.H., T.H.S.,
1029 L.A.R., and M.L.B. collected samples; J.D.D. and L.S. generated the UCE libraries; J.D.D.,
1030 M.E.A., J.C., O.J.L., and P.F.C. analysed and interpreted data; M.F. calibrated tree
1031 reconstructions; J.D.D. wrote the manuscript with input from all co-authors.

Author Manuscript

1032 **Table 1.** Species distribution and clade designation from Bellwood et al. (2010) and Cowman & Bellwood (2011) for all
 1033 Chaetodontidae samples used in this study. Colours in the table header match the colors used to denote species distributions in Figure
 1034 1. Asterisks indicate regional endemics for the purposes of our correlational trait analysis. The letters below each region indicate the
 1035 geographic groupings used for BioGeoBEARS analysis. Although *Chaetodon leucopleura*, *Chaetodon melapturus*, and *Chaetodon*
 1036 *pictus* are listed as being present in the Red Sea, this is based on rare records at their southern limits. Similarly, we have only sampled
 1037 *C. pictus* (and not *Chaetodon vagabundus*) at Socotra (DiBattista et al. 2017), and rare records of *Chaetodon austriacus* in the Gulf of
 1038 Aden and South Oman likely represent waifs.

Species	Geographic distribution								
	Gulf of Aqaba (A)	Rest of Red Sea (B)	Djibouti and Gulf of Aden	Socotra (D)	South Oman (E)	Arabian Gulf (F)	Gulf of Oman and Pakistan	Rest of Indian Ocean	Pacific Ocean (I)
Clade 4									
<i>Chaetodon auriga</i>	√	√	√	√	√		√	√	√
<i>Chaetodon auripes</i>									√
<i>Chaetodon collare</i>			√	√	√		√	√	√
<i>Chaetodon decussatus</i>			√					√	
<i>Chaetodon dialeucos</i> *			√		√				
<i>Chaetodon falcula</i>								√	
<i>Chaetodon fasciatus</i> *	√	√	√						
<i>Chaetodon gardineri</i> *			√	√	√		√	√	
<i>Chaetodon leucopleura</i>		√	√	√	√			√	
<i>Chaetodon lineolatus</i>	√	√	√	√				√	√
<i>Chaetodon lunula</i>			√	√	√			√	√
<i>Chaetodon melannotus</i>	√	√	√	√				√	√
<i>Chaetodon mesoleucos</i> *	√	√	√	√					
<i>Chaetodon nigropunctatus</i> *					√	√	√		
<i>Chaetodon oxycephalus</i>								√	√
<i>Chaetodon pictus</i> *		√	√	√	√		√		
<i>Chaetodon semilarvatus</i> *	√	√	√		√				
<i>Chaetodon vagabundus</i>				√				√	√

Clade 3

<i>Chaetodon austriacus*</i>	√	√	√		√				
<i>Chaetodon baronessa</i>									√
<i>Chaetodon bennetti</i>				√				√	√
<i>Chaetodon larvatus*</i>	√	√	√		√				√
<i>Chaetodon lunulatus</i>									√
<i>Chaetodon melapterus*</i>		√	√	√	√	√	√		√
<i>Chaetodon plebeius</i>									√
<i>Chaetodon spectulum</i>								√	√
<i>Chaetodon triangulum</i>								√	√
<i>Chaetodon trifascialis</i>	√	√	√	√	√			√	√
<i>Chaetodon trifasciatus</i>			√	√				√	√
<i>Chaetodon zanzibariensis</i>			√	√				√	√
<u>Clade 2</u>									
<i>Chaetodon guttatissimus</i>				√				√	√
<i>Chaetodon interruptus</i>								√	√
<i>Chaetodon kleinii</i>				√				√	√
<i>Chaetodon madagaskariensis</i>								√	√
<i>Chaetodon mertensii</i>								√	√
<i>Chaetodon paucifasciatus*</i>	√	√	√						√
<i>Chaetodon pelewensis</i>									√
<i>Chaetodon punctatofasciatus</i>								√	√
<i>Chaetodon trichrous</i>								√	√
<i>Chaetodon unimaculatus</i>								√	√
<i>Chaetodon xanthurus</i>									√
<u>Bannerfishes</u>									
<i>Forcipiger flavissimus</i>		√	√	√				√	√
<i>Forcipiger longirostris</i>				√				√	√
<i>Heniochus acuminatus</i>			√	√	√	√	√	√	√
<i>Heniochus diphreutes</i>	√	√	√	√				√	√
<i>Heniochus intermedius*</i>	√	√	√	√				√	√

1039

1040

1041

1042

1043

1044

1045

Table 2. Akaike information criterion (AIC) model testing based on distribution patterns for butterflyfish lineages using the time-calibrated phylogeny analysed with the R module BioGeoBEARS, where d represents the dispersal parameter, e represents the extinction parameter, and j represents founder-event speciation. For these models we coded each taxon based on presence/absence in nine discrete geographical areas: a) Gulf of Aqaba, b) rest of Red Sea, c) Djibouti and Gulf of Aden, d) Socotra, e) South Oman, f) Arabian Gulf, g) Gulf of Oman and Pakistan, h) rest of Indian Ocean and i) Pacific Ocean. Grey shading indicates the favoured model based on AIC scores.

1052

	Ln Likelihood	Number of parameters	d	e	j	AIC	AIC weight
DEC	-255.13	2	0.06	0	0	514.25	0.03
DEC+J	-250.79	3	0.05	0	0.04	507.58	0.76
DIVALIKE	-253.88	2	0.07	0.04	0	511.76	0.09
DIVALIKE+J	-252.76	3	0.06	0.02	0.03	511.52	0.11
BAYAREALIKE	-259.86	2	0.05	0.18	0	523.71	0
BAYAREALIKE+J	-255.48	3	0.04	0.08	0.06	516.96	0.01

1053

1054

1055

1056

1057

1058

1059

1060

1061

1062

1063

1064

1065

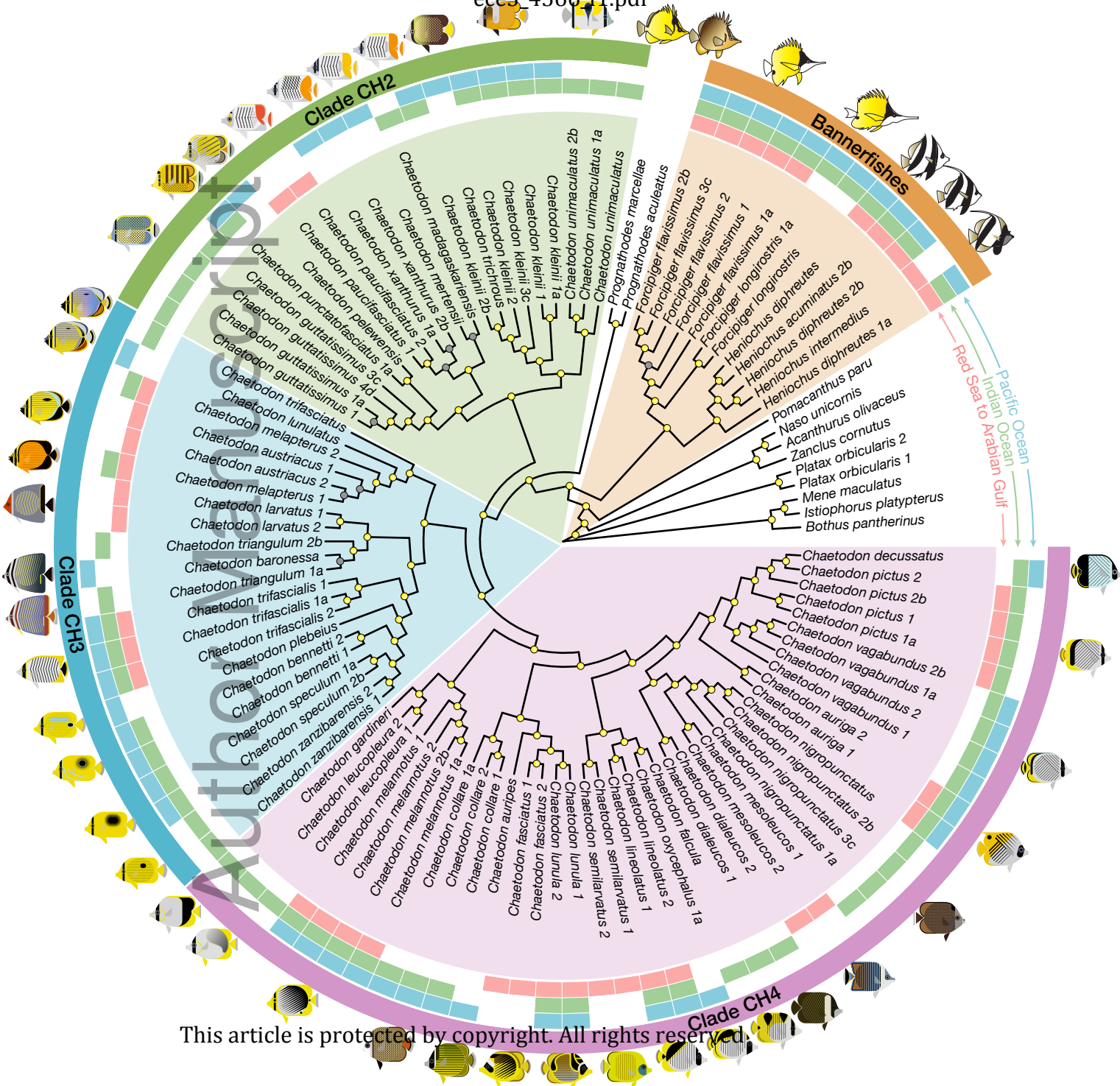
1066

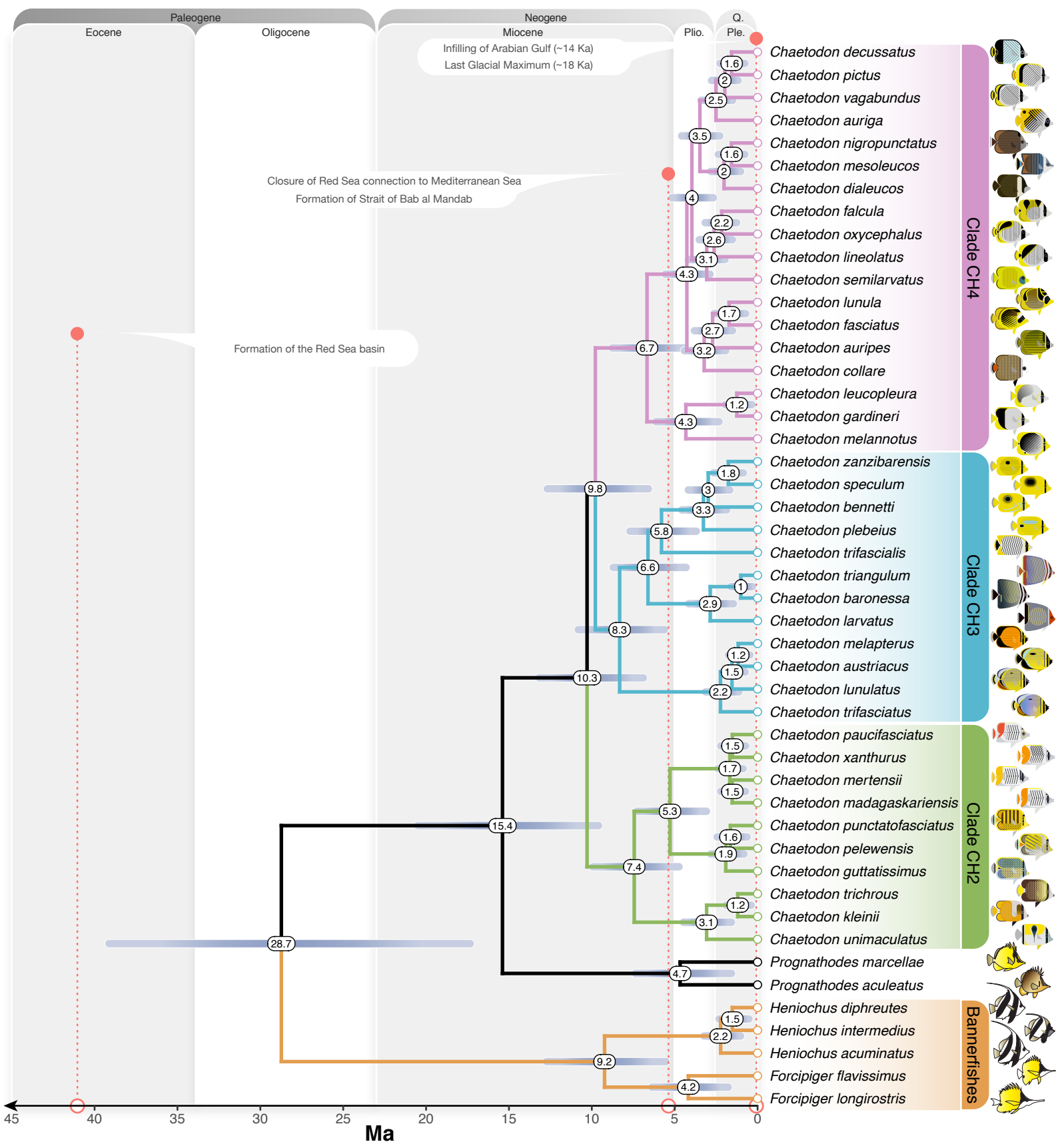
1067

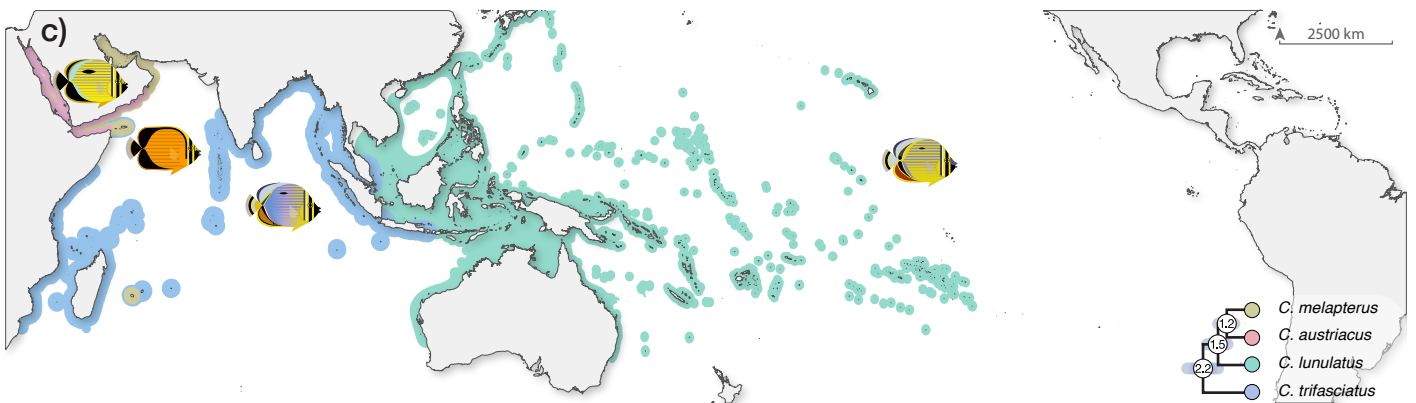
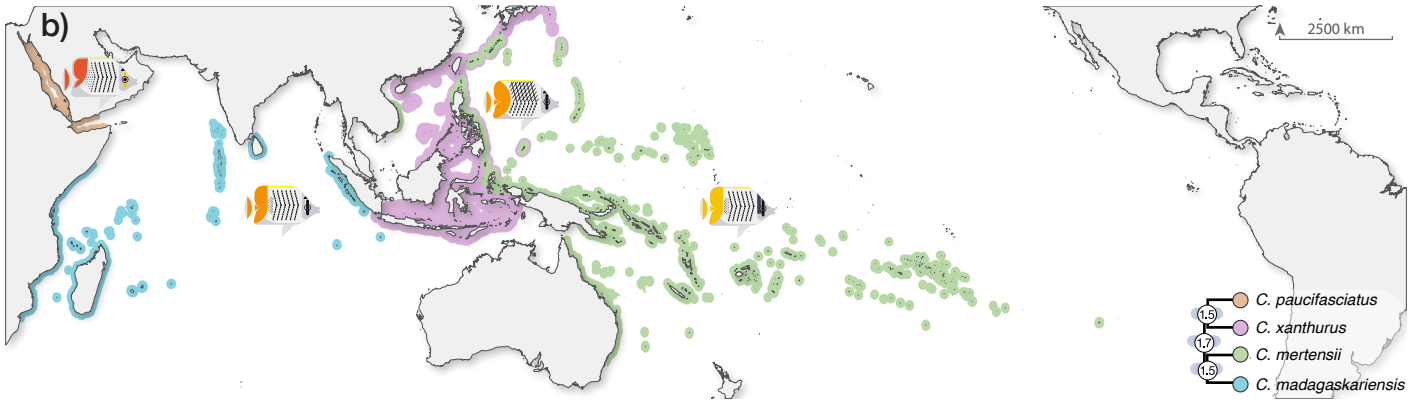
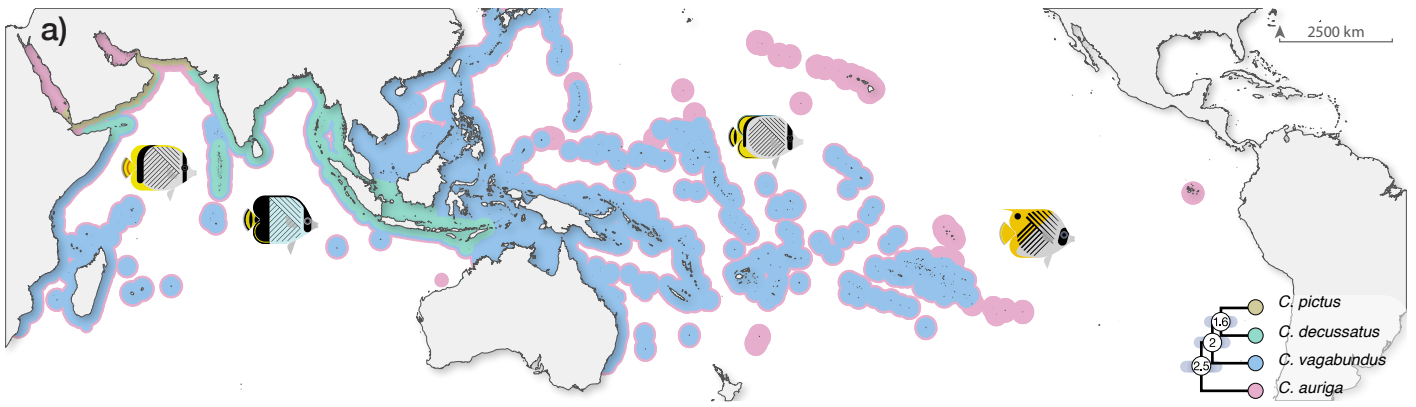
1068 **Table 3.** Summary of the final (best) phylogenetic, linear multi-regression model, based on estimated probability of endemism as a
 1069 response variable, selected after the backward stepwise phylostep procedure. Coefficients in bold indicate significance ($p < 0.05$).

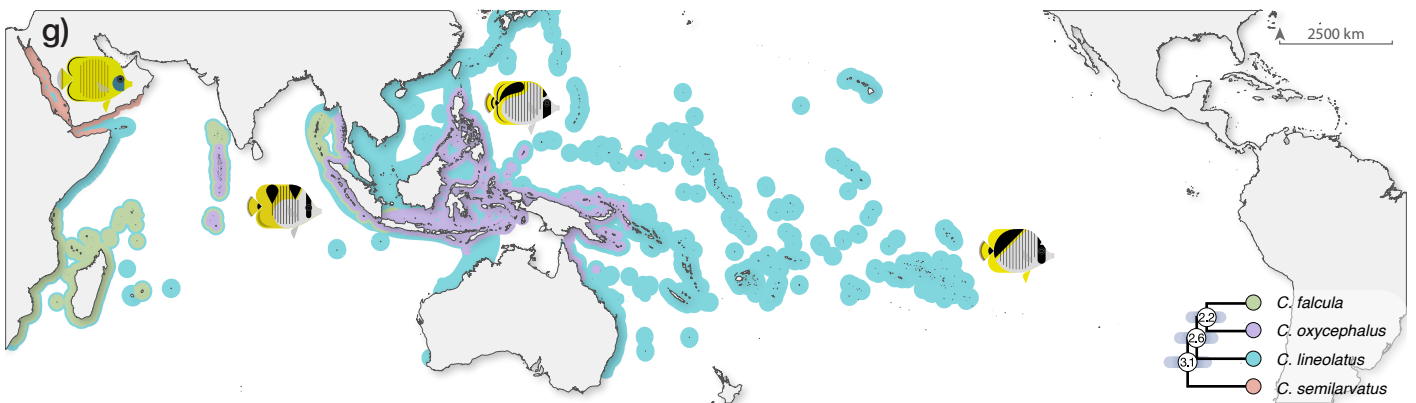
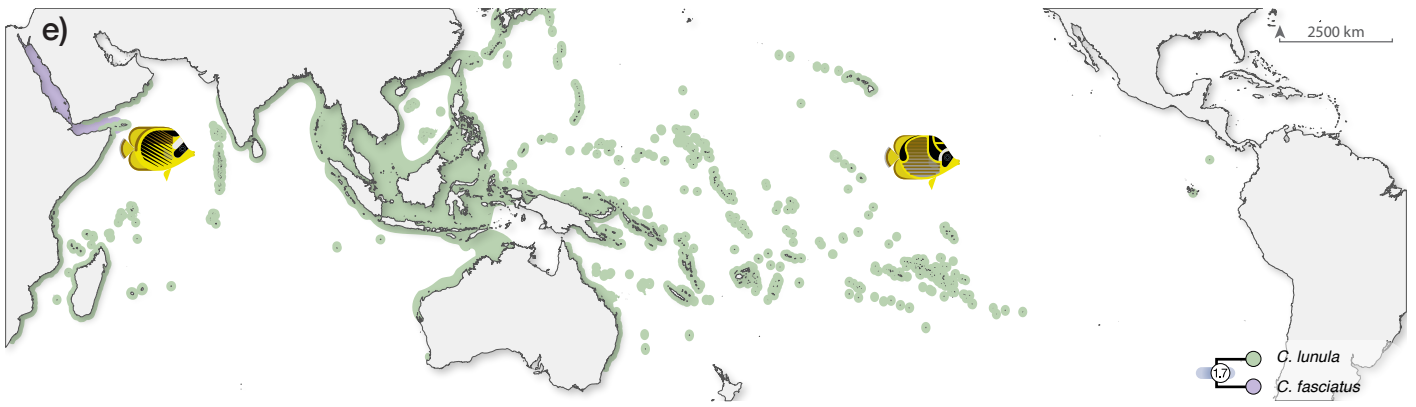
	Estimate	Std. err.	<i>z</i> value	<i>p</i> value
(Intercept)	6.170	2.506	2.461	0.013
Depth range	-1.423	0.543	-2.620	0.008
Phylogenetic age	-1.209	0.694	-1.742	0.061

1070



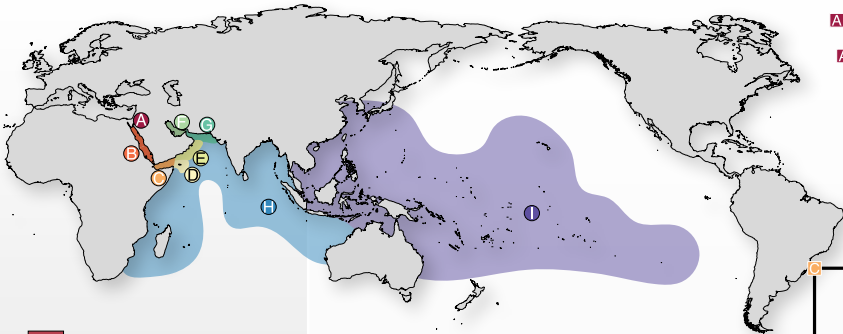




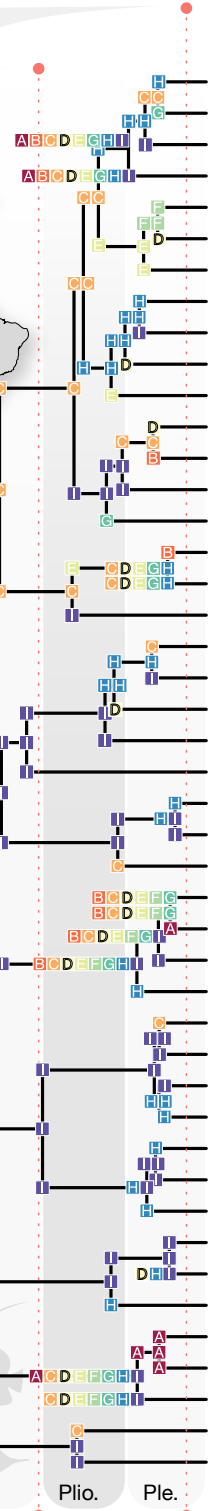
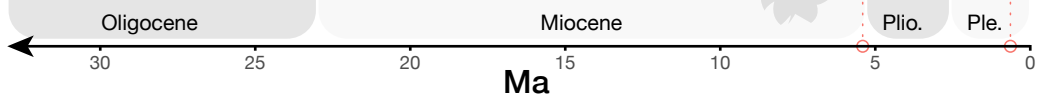


Infilling of Arabian Gulf (~14 Ka)
 Last Glacial Maximum (~18 Ka)

Closure of Red Sea connection to Mediterranean Sea
 Formation of Strait of Bab al Mandab



- A** Gulf of Aqaba
- B** Red Sea
- C** Djibouti & Gulf of Aden
- D** Socotra
- E** South Oman
- F** Arabian Gulf
- G** Gulf of Oman & Pakistan
- H** Indian Ocean
- I** Pacific Ocean



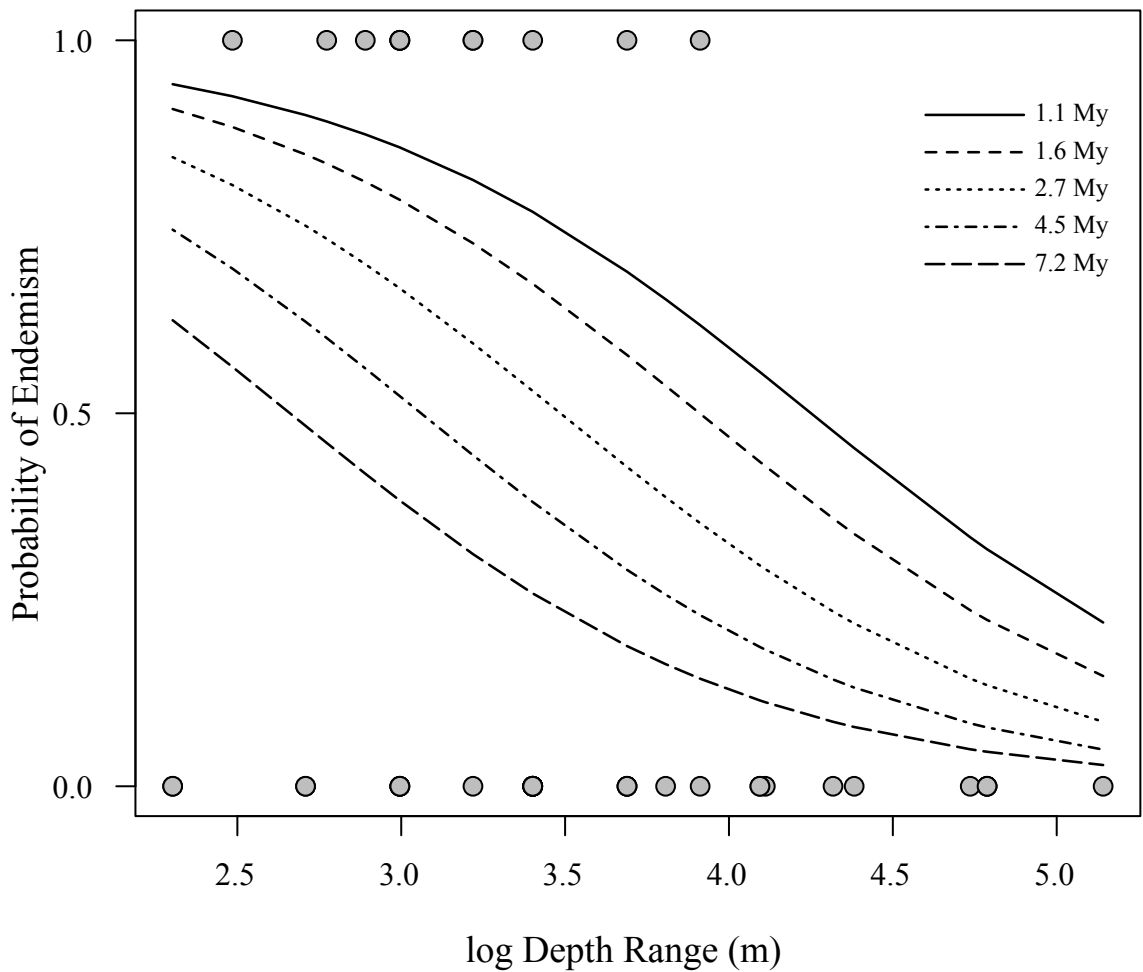
	A	B	C	D	E	F	G	H	I
<i>Chaetodon decussatus</i>									
<i>Chaetodon pictus</i>									
<i>Chaetodon vagabundus</i>									
<i>Chaetodon auriga</i>									
<i>Chaetodon nigropunctatus</i>									
<i>Chaetodon mesoleucos</i>									
<i>Chaetodon dialeucos</i>									
<i>Chaetodon falcula</i>									
<i>Chaetodon oxycephalus</i>									
<i>Chaetodon lineolatus</i>									
<i>Chaetodon semilarvatus</i>									
<i>Chaetodon lunula</i>									
<i>Chaetodon fasciatus</i>									
<i>Chaetodon auripes</i>									
<i>Chaetodon collare</i>									
<i>Chaetodon leucopleura</i>									
<i>Chaetodon gardineri</i>									
<i>Chaetodon melannotus</i>									
<i>Chaetodon zanzibarensis</i>									
<i>Chaetodon speculum</i>									
<i>Chaetodon bennetti</i>									
<i>Chaetodon plebeius</i>									
<i>Chaetodon trifascialis</i>									
<i>Chaetodon triangulum</i>									
<i>Chaetodon baronessa</i>									
<i>Chaetodon larvatus</i>									
<i>Chaetodon melapterus</i>									
<i>Chaetodon austriacus</i>									
<i>Chaetodon lunulatus</i>									
<i>Chaetodon trifasciatus</i>									
<i>Chaetodon paucifasciatus</i>									
<i>Chaetodon xanthurus</i>									
<i>Chaetodon mertensii</i>									
<i>Chaetodon madagaskariensis</i>									
<i>Chaetodon punctatofasciatus</i>									
<i>Chaetodon pelewensis</i>									
<i>Chaetodon guttatissimus</i>									
<i>Chaetodon trichrous</i>									
<i>Chaetodon kleinii</i>									
<i>Chaetodon unimaculatus</i>									
<i>Heniochus diphreutes</i>									
<i>Heniochus intermedius</i>									
<i>Heniochus acuminatus</i>									
<i>Forcipiger flavissimus</i>									
<i>Forcipiger longirostris</i>									

Clade CH4

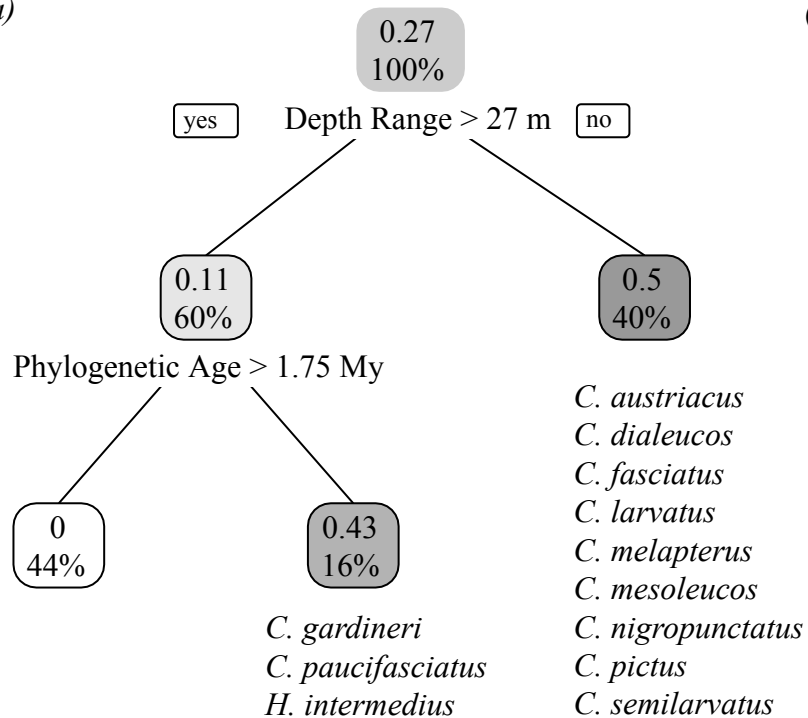
Clade CH3

Clade CH2

Bannerfishes



(a)



(b)

