



HAL
open science

Impact of Pleistocene Eustatic Fluctuations on Evolutionary Dynamics in Southeast Asian Biodiversity Hotspots

Arni Sholihah, Erwan Delrieu-Trottin, Fabien L. Condamine, Daisy Wowor, Lukas Rüber, Laurent Pouyaud, Jean-François Agnèse, Nicolas Hubert

► **To cite this version:**

Arni Sholihah, Erwan Delrieu-Trottin, Fabien L. Condamine, Daisy Wowor, Lukas Rüber, et al.. Impact of Pleistocene Eustatic Fluctuations on Evolutionary Dynamics in Southeast Asian Biodiversity Hotspots. *Systematic Biology*, 2021, 70 (5), pp.940-960. 10.1093/sysbio/syab006 . hal-03370604

HAL Id: hal-03370604

<https://hal.science/hal-03370604>

Submitted on 22 Nov 2021

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

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

1 **Impact of Pleistocene Eustatic Fluctuations on Evolutionary Dynamics in Southeast**
2 **Asian Biodiversity Hotspots**

3

4 Arni Sholihah^{1,2,*}, Erwan Delrieu-Trottin^{2,3}, Fabien L. Condamine², Daisy Wowor⁴, Lukas
5 Rüber^{5,6}, Laurent Pouyaud², Jean-François Agnèsè², Nicolas Hubert²

6

7 ¹*Institut Teknologi Bandung, School of Life Sciences and Technology, Jalan Ganesha 10,*
8 *Bandung 40132, Indonesia.*

9 ²*UMR 5554 ISEM (IRD, UM, CNRS, EPHE), Université de Montpellier, Place Eugène*
10 *Bataillon, 34095 Montpellier cedex 05, France*

11 ³*Museum für Naturkunde, Leibniz-Institut für Evolutions und Biodiversitätsforschung an der*
12 *Humboldt-Universität zu Berlin*

13 ⁴*Division of Zoology, Research Center for Biology, Indonesian Institute of Sciences (LIPI),*
14 *Jalan Raya Jakarta Bogor Km 46, Cibinong 16911, Indonesia.*

15 ⁵*Naturhistorisches Museum Bern, Bernastrasse 15, 3005 Bern, Switzerland*

16 ⁶*Aquatic Ecology and Evolution, Institute of Ecology and Evolution, University of Bern,*
17 *Baltzerstrasse 6, 3012 Bern, Switzerland*

18 **Correspondence to be sent to UMR 5554 ISEM (IRD, UM, CNRS, EPHE), Université de*
19 *Montpellier, Bâtiment 22, Place Eugène Bataillon, 34095 Montpellier cedex 05, France; E-*
20 *mail: arni.sholihah@gmail.com*

21

22 **ABSTRACT**

23 Pleistocene Climatic Fluctuations (PCF) are frequently highlighted as important
24 evolutionary engines that triggered cycles of biome expansion and contraction. While there is
25 ample evidence of the impact of PCF on biodiversity of continental biomes, the consequences

26 in insular systems depend on the geology of the islands and the ecology of the taxa inhabiting
27 them. The idiosyncratic aspects of insular systems are exemplified by the islands of the Sunda
28 Shelf in Southeast Asia (Sundaland), where PCF-induced eustatic fluctuations had complex
29 interactions with the geology of the region, resulting in high species diversity and endemism.
30 Emergent land in Southeast Asia varied drastically with sea level fluctuations during the
31 Pleistocene. Climate-induced fluctuations in sea level caused temporary connections between
32 insular and continental biodiversity hotspots in Southeast Asia. These exposed lands likely
33 had freshwater drainage systems that extended between modern islands: the Paleoriver
34 Hypothesis. Built upon the assumption that aquatic organisms are among the most suitable
35 models to trace ancient river boundaries and fluctuations of landmass coverage, the present
36 study aims to examine the evolutionary consequences of PCF on the dispersal of freshwater
37 biodiversity in Southeast Asia. Time-calibrated phylogenies of DNA-delimited species were
38 inferred for six species-rich freshwater fish genera in Southeast Asia (*Clarias*, *Channa*,
39 *Glyptothorax*, *Hemirhamphodon*, *Dermogenys*, *Nomorhamphus*). The results highlight
40 rampant cryptic diversity and the temporal localization of most speciation events during the
41 Pleistocene, with 88% of speciation events occurring during this period. Diversification
42 analyses indicate that sea level-dependent diversification models poorly account for species
43 proliferation patterns for all clades excepting *Channa*. Ancestral area estimations point to
44 Borneo as the most likely origin for most lineages, with two waves of dispersal to Sumatra
45 and Java during the last 5 Myr. Speciation events are more frequently associated with
46 boundaries of the paleoriver watersheds, with 60%, than islands boundaries, with 40%. In
47 total, one-third of speciation events are inferred to have occurred within paleorivers on a single
48 island, suggesting that habitat heterogeneity and factors other than allopatry between islands
49 substantially affected diversification of Sundaland fishes. Our results suggest that species
50 proliferation in Sundaland is not wholly reliant on Pleistocene sea-level fluctuations isolating

51 populations on different islands. [Dispersal; diversification; eustatic fluctuations; freshwater
52 fishes; insular systems; Milankovitch cycles; paleoenvironments; vicariance]

53

54 INTRODUCTION

55 Pleistocene Climatic Fluctuations (PCF) are widely considered to be drivers of species
56 diversification and distribution patterns (Dodson et al. 1995; Haffer 1997; Lohman et al.
57 2011; de Bruyn et al. 2014) that contributed to the development of highly heterogeneous
58 patterns of species richness and endemism (Voris 2000; Nores 2004; Hubert and Renno 2006;
59 Pellissier et al. 2014). Temporally located between 2.58-0.012 million years ago (Ma)
60 (Walker et al. 2018), PCFs result from dynamic interactions between tectonic movements and
61 oscillations of Earth's orbit, leading to fluctuations of average temperature and global climate.
62 These Milankovitch cycles have impacted biomes through paleoenvironmental dynamics such
63 as eustatic fluctuations. This caused physical fragmentation of terrestrial and freshwater biotas
64 and shaped contemporary distribution patterns (Dodson et al. 1995; Gathorne-Hardy et al.
65 2002; Currie et al. 2004; Bird et al. 2005; Cannon et al. 2009; Beck et al. 2017). Climate
66 differences are frequently invoked to account for heterogeneous speciation and extinction
67 patterns in temperate and tropical biomes. For instance, the Late Pleistocene Hypothesis
68 pinpoints glacial cycles and associated sea-level fluctuations as the main drivers of increased
69 speciation in the Northern Hemisphere (Barraclough and Nee 2001; Wiens and Donoghue
70 2004; Mittelbach et al. 2007; Beck et al. 2017). In the tropics, repeated marine incursions and
71 lowland drying resulted in fragmentation of highlands (Haffer 1997; Nores 1999; Hubert and
72 Renno 2006) and shrinking of terrestrial biotas into refugia, leading to extinction of coastal
73 organisms (Condamine et al. 2015b) or the divergence of isolated populations across refugia
74 (Bates et al. 1998; Nores 1999; Hubert and Renno 2006; Condamine et al. 2015). Pleistocene
75 Climatic Fluctuations are frequently considered to be a “species pump” that offered increased

76 opportunities of speciation through cycles of alternating species' ranges contraction and
77 populations' fragmentation during glacial maxima with dispersal and species' ranges
78 expansion during inter-glacial times (Esselstyn and Brown 2009; April et al. 2012; Brown et
79 al. 2013; Papadopoulou and Knowles 2015a, 2015b; Li and Li 2018). As such, PCFs created
80 opportunities for speciation in otherwise physically stable landscapes in the short-term. Over
81 the long-term, PCFs interacted with geology, resulting in intricate outcomes. Some of these
82 constitute model systems to explore the consequences of PCFs and associated sea-level
83 fluctuations on spatial biodiversity patterns (Weigelt et al. 2016). The Pleistocene Aggregate
84 Island Complex (PAIC) model, for instance, describes the species pump hypothesis on
85 islands, most notably in the Philippines, Caribbean and Mediterranean archipelagos (Esselstyn
86 and Brown 2009; Brown et al. 2013; Papadopoulou and Knowles 2015a, 2015b).
87 Understanding the impact of PCFs on diversity patterns is particularly challenging for the
88 implementation of science-based conservation in the most diverse and threatened biotas
89 globally *i.e.* biodiversity hotspots (Myers et al. 2000; Hoffmann et al. 2010).

90 The biodiversity hotspots in Southeast Asia occur in an area where Pleistocene
91 eustatic fluctuations interact with geology (Voris 2000; Woodruff 2010; Lohman et al. 2011),
92 including the landmasses of Sundaland (Java, Sumatra, Borneo and peninsular Malaysia).
93 Sundaland emerged during the early Cenozoic (ca. 66 Ma) as a promontory at the southern
94 end of Eurasia (Lohman et al. 2011). The isolation of Sundaland, however, is much more
95 recent (Fig. 1). Complex tectonic movements and subduction activities during the Miocene
96 triggered the formation of Borneo between 20 and 10 Ma (Figs. 1a & 1b) and the subsequent
97 emergence of Sumatra and Java between 10 and 5 Ma (Figs. 1b & 1c). From a geological
98 perspective, Sundaland's configuration has been stable for the past 5 Ma (Lohman et al. 2011;
99 Hall 2013). However, insular Sundaland was not completely separated from mainland
100 Southeast Asia until the Pliocene (5.33-2.58 Ma) (Hall 2009; Lohman et al. 2011). Upon

101 entering the early Quaternary (2.58 Ma), sea level dropped from 40 to 60 m below present
102 levels over a ca. 41 kyr period (Lisiecki and Raymo 2005; Miller et al. 2005; Cannon et al.
103 2009; Lohman et al. 2011; de Bruyn et al. 2014) forming land bridges among the islands of
104 Sundaland and the mainland (Fig. 1d, -60 m). Around 800 kyr, Sundaland experienced longer
105 glacial cycles of 100 kyr and lower sea level during eustasy (Lisiecki and Raymo 2005; Miller
106 et al. 2005) which enlarged land bridges and increased fragmentation during interglacial
107 times. During the Last Glacial Maximum (LGM) ca. 17 kyr, sea levels dropped to 120 m
108 below presents (Voris 2000; Sathiamurthy and Voris 2006) (Fig. 1d), resulting in the widest
109 landmass extension to date. At the end of the LGM, rising sea levels shrank terrestrial
110 habitats, and Sundaland biotas are currently considered to be refugial (Cannon et al. 2009;
111 Woodruff 2010; Lohman et al. 2011). Interactions between the geological history of
112 subduction and PCFs are expected to have significantly affected the diversification of
113 terrestrial and freshwater biodiversity on Sundaland (Voris 2000; Sathiamurthy and Voris
114 2006; Lohman et al. 2011; de Bruyn et al. 2013).

115 Eustatic fluctuations in Sundaland not only altered land exposure over time but also
116 impacted the configuration of Sundaland's watersheds. During glacial maxima, sea levels
117 were especially low, allowing riverine watersheds to expand to newly exposed land that was
118 previously sea floor (Voris 2000; Sathiamurthy and Voris 2006; Lohman et al. 2011). These
119 temporary paleoriver systems connected Sundaland's rivers and created temporary linkages
120 between Sundaland and continental Southeast Asia. Particularly well-documented for the
121 Pleistocene LGM, the lower sea level created four large paleoriver systems that connected
122 insular rivers across Sundaland's landmasses (Kottelat et al. 1993; Voris 2000; Woodruff
123 2010) including the paleorivers of East Sunda (1 in Fig. 1d), North Sunda (2 in Fig. 1d),
124 Malacca Straits (3 in Fig. 1d) and Siam (4 in Fig. 1d). For freshwater-dependent organisms,
125 the existence of these paleoriver systems has been long considered to be one of the main

126 drivers of contemporary distribution patterns of freshwater organismal in the region (Kottelat
127 et al. 1993; Voris 2000; Lohman et al. 2011). Formally known as the Paleoriver Hypothesis,
128 the impact of paleorivers on freshwater species distribution has been investigated in
129 *Hemibagrus nemurus* (Dodson et al. 1995), viviparous halfbeaks (de Bruyn et al. 2013),
130 snakeheads (Tan et al. 2012), killifishes (Beck et al. 2017) and *Macrobrachium* shrimps (De
131 Bruyn et al. 2004). These studies detected some congruence between lineages distribution and
132 boundaries of paleoriver watersheds. However, the determinants of these correlations are still
133 debatable due to the limited number of taxa examined (Dodson et al. 1995; Lohman et al.
134 2011; Tan et al. 2012; de Bruyn et al. 2013; Beck et al. 2017). Furthermore, the ages of the
135 taxa in these studies were generally older than the Pleistocene, casting doubt on the role of
136 PCF in freshwater diversification (Dodson et al. 1995; Beck et al. 2017).

137 Here, we explore the evolutionary dynamics of species proliferation and spatial
138 patterns of speciation among Southeast Asian freshwater fishes to address the following
139 questions: (1) Did paleorivers fragment populations of widely distributed taxa? (2) Did
140 paleorivers enable dispersal between islands during glacial maxima? (3) Did PCFs impact the
141 pace of diversification?

142 From this set of initial questions, we derived several predictions made by the
143 Paleoriver Hypothesis regarding the evolution of species range boundaries during speciation
144 (*e.g.*, congruence between species range distribution and boundaries of paleoriver watersheds)
145 and the timing of speciation (*e.g.* Pleistocene species pump) to address the evolutionary
146 response of Southeast Asian freshwater biotas to eustatic fluctuations. To test predictions
147 made by the Paleoriver Hypothesis, we compiled DNA sequences from widely distributed and
148 well-studied freshwater fish lineages with a variety of life history traits (Kottelat et al. 1993).
149 Our molecular phylogenetic dataset of 1,511 individuals was used to reconstruct the
150 phylogenetic relationships and estimate the divergence times of 110 morphological species

151 belonging to four of the most species-rich freshwater fish lineages in Southeast Asia: *Channa*
152 (30 morphological species), *Zenarchopteridae* (17 morphological species), *Glyptothorax* (50
153 morphological species), and *Clarias* (13 morphological species). Southeast Asian freshwater
154 fishes are poorly known taxonomically and cryptic species are common (de Bruyn et al. 2013;
155 Hubert et al. 2015; Lim et al. 2016; Beck et al. 2017; Nurul Farhana et al. 2018). Thus, all
156 macroevolutionary inferences recognize Molecular Operational Taxonomic Units (MOTUs)
157 delineated through DNA-based species delimitation methods (Blaxter et al. 2005; Ruane et al.
158 2014). Distribution ranges were characterized for these MOTUs and further used to explore
159 spatial and temporal patterns of diversification through ancestral state reconstructions and
160 birth-death modeling approaches.

161

162 **MATERIALS AND METHODS**

163 ***Hypothesis Testing, Taxa Selection and Sampling***

164 The Paleoriver Hypothesis (Kottelat et al. 1993; Voris 2000; Woodruff 2010; de
165 Bruyn et al. 2013) makes several predictions about the geography and timing of speciation.
166 First, the paleorivers might be expected to promote allopatric speciation between populations
167 of species that were once widespread or expanding. If so, range boundaries between sister
168 species that diverged during the Pleistocene might be expected to coincide with boundaries of
169 the paleoriver watersheds. Second, paleorivers might have enabled dispersal between different
170 islands, such as South Borneo and Java in the East-Sunda system during periods of low sea
171 level (Fig. 1d). This hypothesis predicts that species assemblages will be most similar within
172 paleoriver drainage basins, even when those basins span two or more present-day islands.
173 Assemblages between paleoriver drainages on the same present-day island are likely to be
174 markedly different in comparison. Finally, the interactions between geological history and
175 PCFs have intensified during the last 5 Ma with the elevation of Sundaland predicting

176 increased diversification. Estimating speciation events in different paleorivers would allow
177 testing of these hypotheses. However, different species interact with their environment in
178 different ways, which will further influence their dispersal and colonization abilities. Thus,
179 dispersal traits might be expected to influence the evolutionary responses of biotas to
180 environmental changes (Barraclough and Nee 2001; Currie et al. 2004; McPeck 2008; Hubert
181 et al. 2015; Burbrink et al. 2016; Liu et al. 2019). To explore the three predictions made by
182 the Paleoriver Hypothesis, we selected taxa conforming to the following requirements: (1)
183 widespread distribution of lineages across Southeast Asia including Sundaland, (2)
184 comprehensive sampling of described species within the taxon, (3) availability of nuclear and
185 mitochondrial sequences, and (4) varying life history traits such as body size and dispersal
186 ability.

187 Four lineages fulfilled these requirements: (1) the snakehead genus *Channa*
188 (Anabantiformes, Channidae) widely distributed in Asia (Conte-Grand et al. 2017), (2) the
189 walking catfish genus *Clarias* (Siluriformes, Clariidae) widely distributed in Asia (Pouyaud et
190 al. 2009), (3) the catfish genus *Glyptothorax* (Siluriformes, Sisoridae) widely distributed in
191 Asia (Kottelat et al. 1993; Jiang et al. 2011) and, (4) the halfbeak genera of Southeast Asia
192 *Dermogenys*, *Hemirhamphodon* and *Nomorhamphus* (Beloniformes, Zenarchopteridae) (de
193 Bruyn et al. 2013; Lim et al. 2016; Nurul Farhana et al. 2018). Both *Clarias* and *Channa* can
194 disperse easily. *Clarias* species, for instance, have a unique accessory air-breathing organ in
195 the upper part of the branchial cavity that allows survival in oxygen-poor water and on land
196 (Munshi 1961; Maina and Maloiy 1986). Meanwhile, *Channa* is a genus of air-breathing,
197 predatory freshwater fishes with an extended swim bladder, paired supra-branchial chambers,
198 and an interior labyrinth organ that enables them to live out of the water for several days.
199 Several species can move on land and burrow into mud during droughts to keep their bodies
200 wet (Kottelat et al. 1993; Berra 2001; Adamson et al. 2010; Serrao et al. 2014; Rüber et al.

201 2019). These capabilities make *Clarias* and *Channa* resilient to environmental variation, and
202 several species are even considered invasive outside their native range (Serrao et al. 2014;
203 Conte-Grand et al. 2017). The other genera have lower dispersal abilities. *Glyptothorax*, the
204 most species-rich genus in the catfish family Sisoridae, is smaller than *Clarias* and *Channa*,
205 and is not equipped with an air-breathing organ (Berra 2001). Distribution of *Glyptothorax*
206 species are mostly fragmented and confined within isolated mountain rapid streams (Ng and
207 Kottelat 2016; Hutama et al. 2017). The healfbeak genera are viviparous with limited larval
208 dispersal abilities (de Bruyn et al. 2013; Nurul Farhana et al. 2018). Although these genera are
209 geographically widespread, their genetic lineages have more restricted distributions (Downing
210 Meisner 2001; Tan and Lim 2013; Lim et al. 2016; Nurul Farhana et al. 2018). Our dataset
211 comprised 2,211 sequences dataset from 1,511 individuals belonging to 110 nominal species
212 and representing four lineages. Outgroups were then selected for each group, ranging from
213 closely related to distantly related taxa following the phylogenetic classification of bony
214 fishes by Betancur-R et al. (2017). Outgroup selection was further refined according to
215 previously published molecular studies for Clariidae (Pouyaud et al. 2009), Sisoridae (Jiang et
216 al. 2011) and Channidae (Conte-Grand et al. 2017).

217

218 ***Genetic Species Delimitation***

219 Genetic studies have identified substantial cryptic diversity in Southeast Asian
220 freshwater fishes (Nguyen et al. 2008; Pouyaud et al. 2009; Hubert et al. 2015, 2019;
221 Dahruddin et al. 2017; Hutama et al. 2017; Nurul Farhana et al. 2018; Sholihah et al. 2020).
222 Cryptic diversity could lead to biases in phylogenetic reconstruction and diversification
223 analyses, particularly due to the overestimation of divergence age estimates in the absence of
224 the closest relatives (Esselstyn et al. 2009; Patel et al. 2011; Ruane et al. 2014). To avoid
225 these shortcomings, genetic species delimitation methods were used to define MOTUs.

226 Several methods of species delineation have been developed, but each has pitfalls. Agreement
227 among different methods suggest robust delimitation (Kekkonen and Hebert 2014; Kekkonen
228 et al. 2015). Thus, we used four methods of species delimitation including a distance-based
229 method with Automatic Barcode Gap Discovery (ABGD, Puillandre et al. 2012), a network-
230 based method with Refined Single Linkage (RESL, Ratnasingham and Hebert 2013), and two
231 tree-based methods including Poisson Tree Processes (PTP, Zhang et al. 2013) and the
232 Generalized Mixed Yule Coalescent (GMYC, Pons et al. 2006). A final delimitation scheme
233 was then established based on a 50% consensus (Hubert et al. 2019). All sequences used were
234 aligned using MUSCLE (Edgar 2004) implemented in MEGA7 (Kumar et al. 2016) and
235 manually edited.

236 These four delimitation methods use different input data. We used sequence
237 alignments to carry out ABGD and RESL analyses. The ABGD analysis was performed
238 through the online platform using the K2P substitution model
239 (<https://bioinfo.mnhn.fr/abi/public/abgd/abgdweb.html>). The RESL algorithm was
240 implemented through BOLD systems version 4 (<http://www.boldsystems.org>, Ratnasingham
241 and Hebert 2007), using Barcode Index Numbers (BIN, Ratnasingham and Hebert 2013). We
242 used maximum likelihood (ML) phylogenetic trees for the PTP and GMYC analyses. For
243 PTP, the analyses were carried out on the web service (<https://mptp.h-its.org/#/tree>). Only the
244 single threshold version was used because preliminary analyses using the multiple threshold
245 version of PTP were too conservative compared to other methods (*i.e.* too few species).
246 Maximum Likelihood phylogenies were reconstructed using RAxML (GUI 1.5) with the
247 GTR+I+ Γ model, and 5000 non-parametric bootstrap (BP) replicates were computed using
248 RAxML-HPC Blackbox (Miller et al. 2010) with RAxML 8 (Stamatakis 2014). For GMYC,
249 the multiple thresholds function (mGMYC) was implemented with the R-package SPLITS
250 1.0-19 (Fujisawa and Barraclough 2013). For GMYC, ultrametric trees were reconstructed

251 using the *chronopl* function in the R-package *ape* 5.3 (Paradis et al. 2004; Paradis 2012;
252 Paradis and Schliep 2019). Time calibration was constrained with the widely accepted fish
253 substitution rate of 1.2% of genetic divergence per million years (Myr) for mitochondrial
254 protein-coding genes (Knowlton et al. 1992) and applied it to the maximum K2P distances,
255 computed using mitochondrial protein coding-genes only, between descendant clades of
256 selected nodes (Table 1). Reference nodes were selected to cover several levels of depth in the
257 phylogenies and to cover a substantial portion of the total species diversity in the entire
258 mitochondrial data set. A total of 13 nominal species were analyzed for *Clarias*, 50 nominal
259 species for *Channa*, 30 nominal species for Zenarchopteridae and 17 nominal species for
260 *Glyptothorax*.

261

262 ***Phylogenetic Reconstruction and Dating***

263 Bayesian phylogenetic reconstructions were implemented with StarBEAST2 package
264 (Ogilvie et al. 2017) of the BEAST2 suite (Heled and Drummond 2010; Bouckaert et al.
265 2014). This approach implements a mixed-model including a coalescent component within
266 species and a diversification component between species that allows accounting for variations
267 of substitution rates within and between species (Ho and Larson 2006). StarBEAST2 requires
268 the designation of species, which were determined using the consensus of our species
269 delimitation analyses. Best substitution models for each marker were estimated using
270 jModelTest2 0.1.10 (Guindon and Gascuel 2003; Darriba et al. 2015). Preliminary analyses
271 conducted on Zenarchopteridae indicated that estimating substitution model parameters
272 jointly with node ages and topologies led to non-converging MCMC with widely fluctuating
273 ages of the Most Recent Common Ancestor (MRCA). Hence, ML parameter estimates were
274 used for MCMC searches in StarBEAST2 with no further estimation of the substitution model
275 parameters. This analytical choice constrained the number of substitution models available in

276 StarBEAST2 and if the selected model was not supported in StarBEAST2, the next most
277 likely model from jModelTest2 available in StarBEAST2 was used. An uncorrelated
278 lognormal clock model was applied alongside estimation of clock rates for all analyses. The
279 reference nodes calibrated for the species delimitation analyses were also used in the
280 Bayesian analyses (Table 1) and further used with a calibrated Yule model. Age calibration
281 was added under normal distribution priors, using a sigma of 1 in order to use wide
282 distributions (Table 1). Yule birthrate value for the analysis was generated using the R-
283 package *ape* by applying a Yule function on each ultrametric tree generated using the
284 *chronopl* function for the GMYC delimitation. To yield enough statistical power (ESS values
285 reaching >200), multiple MCMC runs with 10 million pre-burnin generations each were
286 generated in parallel with tree states stored every 5,000 generations. Considering the number
287 of individuals sampled and molecular heterogeneity within each taxon, the length of the
288 MCMC was different for each group, ranging from 100 million for *Glyptothorax* to 250
289 million for *Channa*. The MCMC were run at least twice to check for the convergence of the
290 chains and reach ESS values > 200 for MRCA age and parameters of the diversification and
291 clock models. The statistics were visualized using Tracer 1.7.1 (Rambaut et al. 2018) to check
292 the outputs and to determine the best converging combination of burnin for each StarBEAST2
293 run. Independent runs were then combined to a single concatenated dataset using
294 LogCombiner 1.10.4 (Bouckaert et al. 2014) to compute the final Bayesian statistics and
295 sampled trees, with or without resampling (depending on the chain lengths). The maximum
296 clade credibility tree, age estimates and corresponding 95% highest posterior density (HPD)
297 were summarized using TreeAnnotator 1.10.4 (Bouckaert et al. 2014).

298 Molecular dating constitutes a crucial step in macroevolutionary inferences
299 (Condamine et al. 2015a) and divergence time estimates can widely vary according to
300 calibration methods and clock models (Ho and Larson 2006; Duchêne et al. 2014). We

301 estimated the impact of clock rates on MOTUs age estimates and diversification trends. For
302 each clade, we rescaled the StarBEAST2 maximum credibility tree according to an
303 assortment of molecular clocks ranging from 0.4% to 1.6% per Ma, with an increment of
304 0.2%, to cover a wide range of known mitochondrial substitution rates (Orti Meyer, A. 1997;
305 Hardman and Lundberg 2006; Read et al. 2006; Kadarusman et al. 2012). StarBEAST2 trees
306 were rescaled using the Alter/Transform Branch Lengths function in Mesquite 3.61
307 (Maddison and Maddison 2019) resulting in a total of seven trees computed for 0.4, 0.6, 0.8,
308 1.0, 1.2, 1.4 and 1.6% per Myr clocks for each group.

309

310 ***Diversification Rates Estimations***

311 We first plotted lineages through time (LTT) for each taxon using the R-package *ape*
312 (Paradis and Schliep 2019) with confidence intervals computed with 1,000 random trees
313 sampled from the StarBEAST2 MCMC file (Heled and Drummond 2010; Bouckaert et al.
314 2014). Then, to test the hypothesis that speciation was linked to past environmental changes
315 (Condamine et al. 2013, 2019), we designed a ML framework including five types of
316 diversification models: constant-rate, time-dependent, temperature-dependent, sea-level-
317 dependent, and diversity-dependent models. These models rely on the ML framework
318 originally developed by Morlon et al. (2011) and implemented in the R-package *RPANDA* 1.3
319 (Morlon et al. 2016). We accounted for missing species in the phylogeny by using a global
320 sampling fraction that is the ratio of sampled species diversity over the total described species
321 diversity for each clade. We designed and fit a total of 17 diversification models (Table S1).
322 We first fit two constant-rate models, considered as references for model comparison: one
323 with the speciation rate constant through time with no extinction (BCST) and another with
324 speciation and extinction rates constant through time (BCSTDCST). Second, we fit four time-
325 dependent models: a model with only speciation rate varying through time and no extinction

326 (BtimeVar), a model with speciation rate varying through time and constant extinction
327 (BtimeVarDCST), a model with constant speciation and extinction rate varying through time
328 (BCSTDtimeVar), and a model with speciation rate and extinction rate both varying through
329 time (BtimeVarDtimeVar). We then fit eight models with speciation and extinction rates
330 varying according to an external environmental variable of which four models had the
331 temperature curve, and four had the sea level curve as follows: two models with speciation
332 varying in function of the environmental variable (BtemperatureVar and Bsea-levelVar), two
333 models with speciation varying in function of the variable with constant extinction rate
334 (BtemperatureVarDCST and Bsea-levelVarDCST), four models with extinction rate varying
335 as a function of the variable and constant speciation rate (BCSTDtemperatureVar and
336 BCSTDsea-levelVar), and two models with both speciation and extinction rates varying as a
337 function of the variable (BtemperatureVarDtemperatureVar and Bsea-levelVarDsea-
338 levelVar). All diversification analyses were performed independently for each the four fish
339 clades.

340 We chose an exponential dependence with time, temperature or sea level, and
341 diversification rates because this exponential function is robust and can take a broad range of
342 shapes depending on the strength and direction of the dependence to the fitted variable.
343 Speciation rate (λ) and extinction rate (μ) are parameterized by the set of following
344 parameters: When speciation and extinction rates are exponential functions of the sea level
345 through time, we used the equations $\lambda(S_{(t)}) = \lambda_0 \times e^{\alpha S_{(t)}}$ and $\mu(S_{(t)}) = \mu_0 \times e^{\beta S_{(t)}}$, where λ_0
346 and μ_0 are respectively the speciation rate and the extinction rate expected when sea level is at
347 0 meters. The variables α (and β) are coefficients that measure the strength and sign of the
348 relationship with sea level (e.g. $\alpha > 0$ and $\beta > 0$ indicate that speciation and extinction rates
349 increase with sea level high stands). Similar equations can be written for an exponential
350 relationship between temperature through time and speciation and extinction rates: $\lambda(T_{(t)}) =$

351 $\lambda_0 + \alpha T(t)$ or $\lambda(T(t)) = \lambda_0 \times e^{\alpha T(t)}$, where $T(t)$ is the temperature at time t and λ_0 is the
352 speciation rate at the temperature of 0° C.

353 In addition, three diversity-dependent diversification models were fitted using a
354 maximum likelihood function with the R-package *DDD* 3.7 (Etienne et al. 2012). In the
355 diversity-dependent models, speciation rates or extinction rates vary as a function of the
356 number of lineages in the clade (Etienne et al. 2012). We took this function to be linear as
357 explained in Etienne et al. (2012). The diversity-dependent models are parameterized by λ_0
358 and μ_0 , the speciation and extinction rates in the absence of a competing lineage, and K that is
359 the ‘carrying capacity’, and represents asymptotic clade size. All speciation and extinction
360 rates were constrained to be positive. The models were compared with the corrected Akaike
361 Information Criterion (AICc) and Akaike weight (AIC ω). The model with the lowest AICc
362 and highest AIC ω was considered to be the best fitting model for the phylogeny. This model
363 selection analysis was repeated for each of the seven dated trees based on different clock rate
364 hypotheses (0.4% to 1.6% per Ma) for the groups departing from a constant diversification
365 model and related to sea level or temperature.

366

367 *Ancestral States Estimations*

368 To infer the influence of Sundaland paleoriver and island systems on the
369 diversification of Southeast Asian freshwater fishes, we estimated the ancestral areas of origin
370 for each clade using the species trees obtained from StarBEAST2. Ancestral area estimations
371 were performed with the R-package *BioGeoBEARS* 1.1 (Matzke 2013) using two sets of
372 geographic delimitations based on: (1) paleorivers, and (2) islands. Then, geographic patterns
373 of speciation were recorded as follows: (1) no dispersal, sister species co-occur within the
374 same paleoriver and the same island; (2) dispersal between islands within a paleoriver
375 forming sister species on different islands within the same paleoriver; (3) dispersal between

376 paleorivers within the same island forming sister species on different paleorivers within the
377 same islands; and (4) dispersal between islands and between paleorivers forming sister
378 species on different paleorivers and different islands. Ancestral area estimations involving
379 paleorivers were based on the following geographic areas (Fig. 1): (1) the North Sunda river
380 system, (2) the East Sunda river system, (3) the Malacca Straits river system, (4) the Siam
381 river system, (5) the Bangka-Belitung, (6) the China-South Asia, (7) the Mekong and
382 Irrawaddy river system, (8) the Northern Borneo and (9) the Philippines. Ancestral area
383 estimations involving islands were based on the following areas: (1) Borneo, (2) Java-Bali,
384 (3) Sumatra, (4) Bangka-Belitung, (5) China-South Asia, (6) Mainland Southeast Asia, (7)
385 Philippines, and (8) Sulawesi. The occurrence of each MOTU in the areas were compiled in a
386 presence/absence matrix for each taxon and served as input data together with the species
387 tree. The range distribution of each MOTU from both approaches is provided in Table S2.
388 Ancestral area estimations were then carried out using 6 models: Dispersal-Extinction
389 Cladogenesis (DEC) and DEC+J (Matzke 2014); ML version of Dispersal-Vicariance
390 analysis (DIVALIKE) and DIVALIKE+J; Bayesian biogeographical inference model
391 (BAYAREALIKE) and BAYAREALIKE+J (Van Dam and Matzke 2016). The inclusion of
392 the parameter J has been recently criticized from a conceptual and statistical perspective (Ree
393 and Sanmartín 2018). The formalization of founder-effect speciation as “jump dispersal” has
394 been developed for insular systems to account for divergence of a new lineage in a new
395 geographic area established by colonization without an intermediate widespread ancestor
396 (Clark et al. 2008; Ree and Sanmartín 2018). Considering the biogeographic scenario of
397 Sundaland and the insularity of the system, jumping dispersal cannot be discarded *a priori*
398 from a conceptual perspective and several studies have previously highlighted the importance
399 of founder-effect speciation in insular systems (de Bruyn et al. 2013; Condamine et al. 2015b;
400 Beck et al. 2017). In fact, the Paleoriver Hypothesis predicts that sea level low stands enabled

401 jump dispersal among islands through temporary land bridges. Models of ancestral area
402 estimation including the J parameters were thus considered here and the best-fit model was
403 estimated using the AICc.

404

405 RESULTS

406 *Phylogenetic Reconstruction and Species Delimitation*

407 A total of 2,211 sequences from 1,511 individuals were mined from *GenBank* and
408 BOLD, representing 110 nominal species (Table S2). The alignment for the genus *Clarias*
409 consisted of 5,322 base pairs (bp), including 3,093 bp from the mitochondrial genome (16S,
410 COI and Cytb) and 2,229 bp from the nuclear genome (RAG1, RAG2), for 146 individuals
411 belonging to 13 morphological species. The alignment for the genus *Channa* comprised of
412 2,856 bp, including 2,076 bp from the mitochondrial genome (16S, COI, Cytb) and 780 bp
413 from the nuclear genome (RAG1), for 891 individuals belonging to 30 morphological species.
414 The alignment for the genera *Dermogenys*, *Hemirhamphodon* and *Nomorhamphus* consisted
415 of 4,109 bp, including 984 bp from the mitochondrial genome (COI, Control Region) and
416 3125 bp from the nuclear genomes (DP5, DP14, DP21, DP35, DP37, HP5, HP54, HPR56),
417 for 266 individuals belonging to 17 morphological species. The alignment for the genus
418 *Glyptothorax* data consisted of 2,776 bp, including 1,869 bp for the mitochondrial genome
419 (COI, Cytb) and 907 bp for the nuclear genome (RAG2), for 208 individuals belonging to 50
420 morphological species. All alignments are available in TreeBASE (TB2:S26912).

421 Substitution models for each marker are provided in Table S3 (Online supplementary
422 material). The four ML phylogenetic reconstructions generally provided well-supported
423 clades with most internal nodes supported by BP > 80 except within the genus *Channa* (Fig.
424 2, Fig. S1). For the genus *Clarias*, Asian species constitute a monophyletic group (Fig. 2A,
425 clades II + III) separated from the African species (Fig. 2A, clade I), with *Clarias gariepinus*

426 (Burchell 1822) as sister to all remaining species. All species are monophyletic except *C.*
427 *nieuhofii* Valenciennes 1840, which is recovered as polyphyletic with three distinct lineages
428 (Fig. S1). For the genus *Glyptothorax*, the inferred tree is imbalanced with continental species
429 constituting a stem group at the root of the tree (Fig. 2B, Fig. S1) and Sundaland species
430 constituting a monophyletic group (Fig. 2B, clade I). All species are monophyletic, except *G.*
431 *platypogonides* (Bleeker 1855) and *G. fuscus* Fowler 1934, which are polyphyletic (Fig. S1).
432 The monophyly of the Zenarchopterid genera *Hemirhamphodon* (Fig. 2C, clade I)
433 *Nomorhamphus* (Fig. 2C, clade II), and *Dermogenys* (Fig. 2C, clade III) is recovered with a
434 sister-relationship between *Dermogenys* and *Nomorhamphus* (Fig. 2C, clades II + III). All
435 species are monophyletic, except *Nomorhamphus megarrhamphus* (Brembach 1982), which
436 is paraphyletic (Fig. S1). For the genus *Channa*, internal relationships are poorly supported,
437 but four main clades are observed (Fig. 2D, clades I to IV). All *Channa* species are
438 monophyletic, except *Channa gachua* (Hamilton 1822), which is polyphyletic and includes at
439 least three lineages distinctly associated to other species, and *C. lucius* (Cuvier 1831) and *C.*
440 *marulius* (Hamilton 1822), which are paraphyletic (Fig. S1).

441 As expected, different species delimitation analyses yielded variable number of
442 MOTUs (Table S2): (1) 47 *Clarias* using mGMYC, 32 using PTP, 7 using RESL and 24
443 using ABGD, resulting in a consensus of 29 MOTUs, (2) 107 *Glyptothorax* using mGMYC,
444 58 using PTP, 58 using RESL and 72 using ABGD, resulting in a consensus of 61 MOTUs,
445 (3) 48 Zenarchopteridae using mGMYC, 27 using PTP, 35 using RESL and 49 using ABGD,
446 resulting in a consensus of 43 MOTUs, and (4) 521 *Channa* using mGMYC, 120 OTUs using
447 PTP, 57 using RESL and 133 using ABGD resulting in a consensus of 120 MOTUs. The
448 consensus delimitation scheme resulted in between 1.22 (*Glyptothorax*) and 4.00 (*Channa*)
449 more MOTUs than morphological species (Table S2).

450 The final Bayesian analyses as implemented in StarBEAST2 were performed using an
451 MCMC of length required to reach convergence (ESS > 200). For *Clarias*, two parallel runs
452 were run with a burnin of 65.01 million generations and 15 million generations, respectively,
453 and both runs were concatenated into a 220-million-generation dataset of 44,000 sampled
454 trees. For *Glyptothorax*, three parallel runs were launched with burnins of 5, 10 and 5 million
455 generations, respectively, and then concatenated into a 280-million-generation dataset with
456 56,000 sampled trees. For *Channa*, four parallel runs were used with burnins of 182.5, 27.52,
457 47 and 43 million generations, respectively, and concatenated into a 700-million-generation
458 dataset with 35,000 sampled trees. For Zenarchopteridae, four parallel runs were used with
459 burnin of 70, 50, 38 and 60 million generations, respectively, and concatenated into a 182-
460 million-generation dataset with 18,200 sampled trees. The StarBEAST2 chronograms were
461 largely congruent with the topologies estimated with the ML approach (Fig. S2). The ages of
462 the MRCA vary between clades with a MRCA dated at 7.09 Ma (95% HPD = 4.49-10.31
463 Ma,) for *Clarias*, 7.1 Ma (95% HPD = 4.51-9.54 Ma) for *Glyptothorax*, 12.63 Ma (95% HPD
464 = 8.24-18.68 Ma) for Zenarchopteridae, and 14.35 Ma (95% HPD = 10.54-18.41 Ma) for
465 *Channa* (Fig S2). For *Clarias*, the MRCA of the Asian species is dated around 6.37 Ma (95%
466 HPD = 3.99-9.19 Ma) and the Zenarchopteridae genera are dated around 9.88 Ma (95% HPD
467 = 6.02-14.70 Ma) for *Hemirhamphodon*, 5.88 Ma (95% HPD = 4.02-8.09 Ma) for
468 *Dermogenys*, and 3.90 Ma (95% HPD = 2.14-5.94 Ma) for *Nomorhamphus*.

469

470 ***Diversification Rates***

471 Most of the MOTUs delineated in the consensus scheme originated in the Pleistocene
472 with divergence time estimates younger than 2.5 Myr (Fig. 3). More than half (53%) of the
473 morphological species and 88% MOTUs are younger than 2.5 Myr (Fig. 3A). The LTT plots
474 indicate that 84% of extant fish diversity occurred during the Pleistocene. On Sundaland,

475 Pleistocene speciation peaks during sea level low stands with 88% of speciation events
476 inferred during the last 2.5 million years and 76% during the last 1.5 million years. This peak
477 of Pleistocene speciation is observed for a wide range of clock rates, including some of the
478 slowest clock rate hypothesis with 72% and 57% of speciation events during the last 2.5 and
479 1.5 Myr, respectively, for the 0.6% per million years (Fig. S3). Faster rates yielded a higher
480 percentage of young MOTUs. The 1.6% per Myr rate yielded 93% and 83% of MOTUs
481 younger than 2.5 and 1.5 Myr, respectively. The only exception was observed for the 0.4%
482 per million years with 60% and 38% of the MOTUs younger than 2.5 and 1.5 Myr,
483 respectively.

484 Likelihood scores and AIC_{ω} for 17 diversification models indicate clade-specific
485 patterns of diversification (Table S4). For *Clarias* and *Glyptothorax*, the constant-rate
486 speciation model is the most likely according to AIC_{ω} , respectively 0.267 and 0.212, with
487 high speciation rates (λ) of 0.4556 and 0.5444, respectively. For Zenarchopteridae, three
488 models are equally likely, including: (1) the time-dependent speciation model ($AIC_{\omega} = 0.145$,
489 $\lambda_0 = 0.432$) with a speciation increasing through time ($\alpha = -0.0971$); (2) the constant-rate
490 speciation model ($AIC_{\omega} = 0.119$, $\lambda_0 = 0.3316$); and (3) the temperature-dependent speciation
491 model ($AIC_{\omega} = 0.116$, $\lambda_0 = 0.5132$) with speciation increasing as temperatures cooled ($\alpha = -$
492 0.1444). While diversification patterns of the previous three groups show no dependency on
493 sea-level eustasy, the most likely model for *Channa* diversification through time is the sea-
494 level-dependent speciation model ($AIC_{\omega} = 0.415$, $\lambda_0 = 0.2279$) with speciation increasing as
495 sea level dropped ($\alpha = -0.0182$). These clade-dependent diversification patterns are reflected
496 by the heterogeneous trends of speciation rates through time for the four lineages (Fig. 3B),
497 including constant speciation rates through time for *Glyptothorax* and *Clarias*, a sea-level
498 dependent speciation rate for *Channa*, and a time-dependent speciation rate for
499 Zenarchopteridae. The dependency of *Channa* diversification on sea level eustasy was

500 recovered for all clock rates except for the two slowest clocks (Table S5). The most likely
501 model for the 0.4% per Myr rate includes a positive relationship between speciation and sea
502 level, speciation increased as sea level increased ($\alpha = 0.0119$). The most likely model for the
503 0.6% per Myr clock includes a positive relationship between extinction rates and sea levels
504 eustasy, extinction increased when sea level increased ($\beta = 0.0351$).

505

506 ***Geography of Diversification***

507 Ancestral area estimations with models incorporating the J parameter were more likely
508 for all groups on islands or paleorivers than models lacking that parameter (Table S6). The
509 DEC+J model is the most likely and was further used in subsequent analyses. Most
510 colonization was initiated before the Pleistocene and in Borneo (the North Sunda and East
511 Sunda paleoriver systems, Fig. 4), while several subsequent colonizations happened through
512 Sumatra, and Java is typically the last island colonized. This pattern was observed within
513 *Clarias*, with a first vicariance event inferred between Borneo and the Southeast Asian
514 mainland at 6.37 Ma (95% HPD = 3.99-9.19 Ma) and associated with the first split in the
515 Asian *Clarias* (Fig. 4A, clade II vs clade III), the colonization of Sumatra happening
516 subsequently twice from Borneo (Fig. 4A, clade III). The same pattern is observed for
517 *Glyptothorax* with a vicariance inferred between Borneo and mainland around 3.64 Ma (95%
518 HPD = 2.34-5.17 Ma; Fig. 4B, clade I). Different patterns are observed for Zenarchopteridae
519 (Fig. 4C) and *Channa* (Fig. 4D), indicating that the islands of Sundaland were colonized
520 multiple times.

521 Most speciation events are inferred to happen within islands, as exemplified by the
522 diversification of three of the four lineages (*Clarias*, *Glyptothorax* and Zenarchopteridae) in
523 Borneo and Sumatra (Figs. 4A, 4B and 4C) or Java for *Channa* (Fig. 4D). The same trend is
524 observed for paleorivers; most speciation events are inferred to occur within them (Fig. 4).

525 Four major geographic patterns of speciation were identified: (1) sister species occupy the
526 same paleoriver on the same island, *e.g.* *Glyptothorax platypogon* 1-4 from Java (East Sunda
527 River System), (2) sister species occupy the same paleoriver on different islands, *e.g.*
528 *Glyptothorax robustus* (Boeseman 1966) 1-4 from Java and Sumatra (East Sunda River
529 System), (3) sister species occupy different paleorivers within the same island, as in *G.*
530 *amnestus* (North Sunda) and *G. fuscus* 4 (Malacca Strait), and (4) sister species occupy
531 different islands and different paleorivers *e.g.* *Clarias pseudoleiacanthus* Sudarto, Teugels
532 and Pouyaud 2003 (North Sunda in Borneo) and *C. leiacanthus* Bleeker 1851 (Malacca Strait
533 in Sumatra).

534 In total, 66.3 % of the inferred speciation events involve dispersal either between
535 islands or between paleorivers (Table 2). In Sundaland, 59.2% of speciation events occurred
536 within island (40.8% between islands) while 39.8% of speciation events occurred within
537 paleorivers (60.2% between paleorivers). These trends are observed for all groups except
538 Zenarchopteridae, in which speciation events within islands represent 81.7% and speciation
539 within paleorivers 63.6% (Table 2). The proportion of speciation events within or between
540 islands is stable through time for speciation resulting from dispersal between paleorivers (Fig.
541 5A). However, this pattern was not observed for Zenarchopteridae (Fig. 5B), where speciation
542 was primarily between islands. In *Glyptothorax*, most speciation events between paleorivers
543 involve dispersal between islands (Fig. 5E). Within paleorivers, speciation events involving
544 dispersal between islands are scarce for all groups (Figs. 5 B, C, D, E).

545

546 **DISCUSSION**

547 ***Diversification of Southeast Asian Aquatic Biotas***

548 The "Late Pleistocene Hypothesis" (Barracough and Nee 2001; Wiens and Donoghue
549 2004; Mittelbach et al. 2007; Beck et al. 2017) and the "Pleistocene Species Pump

550 Hypothesis" (Esselstyn and Brown 2009; Brown et al. 2013; Papadopoulou and Knowles
551 2015a, 2015b; Li and Li 2018) suggest that sea-level fluctuations accelerated diversification
552 of insular biotas. However, we found poor support for these hypotheses and signature of sea-
553 level fluctuations on diversification was apparent in only one of the four fish groups studied
554 here. The genus *Channa* is the only group in which diversification rate was associated with
555 sea-level fluctuations: its speciation rate increased as sea level dropped. This translates into
556 increasing speciation rates through time since sea levels generally decreased through the
557 Pleistocene (Miller et al. 2005). Hence, the genus *Channa* had the highest speciation rate of
558 all studied taxa (Fig. 3B). The superior dispersal ability and environmental adaptability of
559 *Channa* compared to the other taxa may enable them to colonize many different types of
560 environments. With the closure of dispersal routes between islands during sea level
561 highstands, gene flow was reduced, thereby increasing the probability of divergence between
562 populations, eventually leading to speciation. Such speciation events likely occurred
563 repeatedly over short time intervals (glaciation periods). Hence, the genus *Channa* is the only
564 taxon we studied that supports the "Pleistocene Species Pump" hypothesis.

565 On the contrary, the diversification of the other clades does not support this
566 hypothesis. Both silurid groups (*Clarias* and *Glyptothorax*) diversified at a constant rate
567 despite different life history strategies. *Glyptothorax* had a higher speciation rate than *Clarias*,
568 which might be related to the tendency of *Glyptothorax* to occupy hilly streams that are more
569 isolated (Hutama et al. 2017). The dispersal abilities of *Clarias* likely limited allopatric
570 speciation. Such habitat specificity for *Glyptothorax* lineages could have (1) inhibited inter-
571 island dispersal during low seastand since the freshwater corridors between islands were
572 unsuitable, as well as (2) facilitated isolation and *in situ* diversification due to restricted gene
573 flow (Hubert et al. 2015; Hutama et al. 2017), hence generating high, constant speciation. In

574 the same way, none of the diversification trends for Zenarchopteridae were directly related to
575 sea-level fluctuations.

576 Diversification trends show no tendency to decline following a diversity-dependent
577 diversification (DDD) model (Seehausen 2007; Rabosky and Lovette 2008; Seehausen et al.
578 2008). The DDD model is based on the assumption that biotas are bounded by the carrying
579 capacities of the environment, either in terms of habitat diversity (Kisel et al., 2011;
580 Phillimore & Price, 2008; Schluter, 2000) or the total number of individuals that can be
581 sustained (Alonso et al., 2006; Hubbell, 2001). This model predicts that diversification rates
582 decelerate through time as a consequence of (1) declining speciation rates due to the
583 increasing occupation of available niches by proliferating species (Schluter 2000; Gavrilets
584 and Vose 2006; Hubert et al. 2015), or (2) increasing extinction rate through time over
585 constant speciation rate as a consequence of the increasing importance of stochastic
586 demographic dynamics within proliferating species, which are presumed to have reduced
587 population sizes compared to their parental species (Alonso et al. 2006; Rabosky and Lovette
588 2008; Hubert et al. 2015). Here, however, we found no trend for declining diversification
589 rates in any taxa. Two of the four clades have rates that increase through time (Fig. 3B). The
590 high diversity of freshwater habitats in Sundaland likely provides a diversity of niches to
591 sustain elevated and/or constant diversification. Water chemistry is widely diverse in tropical
592 systems (Guyot 1993) and Sundaland is no exception, with water types ranging from clear
593 and alkaline waters in karsts (Clements et al. 2006) and turbid sediment-rich waters of the
594 floodplains to the humic acid-rich waters of the peat swamps (Ng et al. 1994; Giam et al.
595 2012). The large number of fish MOTUs distributed among multiple paleorivers and sister
596 MOTUs diversifying within the same paleoriver seems consistent with the existence of such
597 high diversity of habitat/niche diversity within paleorivers. However, documented examples
598 of ecological speciation within freshwater rivers are scarce (Sullivan et al. 2002; Nolte et al.

599 2005; Alter et al. 2017). Most cases have been discovered in ancient lakes or other lentic
600 water bodies (Nagel and Schluter 1998; Verheyen et al. 2003; Barluenga and Meyer 2004;
601 Herder et al. 2006; Landry et al. 2007). Zenarchopteridae and *Clarias* species are distributed
602 across a wide range of aquatic habitats, with some *Clarias* and *Hemirhamphodon* species
603 occupying acidic waters of peat swamps while other *Clarias* and *Dermogenys* species occupy
604 clear waters and floodplains (Pouyaud et al. 2009; Lim et al. 2016; Nurul Farhana et al. 2018).
605 However, inspection of the phylogenies reconstructed here suggests that transitions between
606 different types of aquatic habitats are scarce; peat swamp and floodplain-associated species of
607 *Clarias* grouped in distinct clades. This trend suggests that adaptive divergence across
608 freshwater habitats is limited in the groups studied here. By contrast, landscape fragmentation
609 by geology has been widely document in the area (Nguyen et al. 2008; Lim et al. 2016;
610 Dahrudin et al. 2017; Hutama et al. 2017; Nurul Farhana et al. 2018; Hubert et al. 2019).

611 Habitat diversity within paleorivers might be expected to have interacted with the
612 availability of dispersal routes among paleorivers, which was affected by sea-level
613 fluctuations (Brown et al. 2013). Similar trends have also been reported for the Pleistocene
614 Aggregate Island Complex (PAIC) model in the Philippines archipelago. Brown et al (Brown
615 et al. 2013) suggested an alternative "nested PAIC model" considering the interplay of the
616 external sea level fluctuations causing cycles of island connection-isolation and changing
617 local habitat diversity within an island. Due to both its paleoriver and archipelagic terrain,
618 Sundaland is likely comparable to the Philippines with extensive and heterogeneous
619 freshwater habitats (Brown et al., 2013; Esselstyn & Brown, 2009; Papadopoulou & Knowles,
620 2015b, 2015a). Frequent regional intermediate disturbance associated with PCF likely
621 impacted biotic interactions among Southeast Asian freshwater fish species and probably
622 triggered a reshuffling of their ecological communities, which may explain the higher species
623 richness than expected under a DDD model.

624

625 ***Diversification Mostly Occurred Within Islands and Paleorivers***

626 Our ancestral area estimations demonstrate that diversification frequently resulted
627 from long-distance dispersal followed by *in situ* diversification (founder-effect speciation) as
628 suggested by the high likelihood of the models with the J parameter (Matzke 2014). This
629 pattern has been previously proposed through detection of trans-island dispersal of freshwater
630 fishes (Dodson et al. 1995; Pouyaud et al. 2009; Adamson et al. 2010; Tan et al. 2012; Tan
631 and Lim 2013; de Bruyn et al. 2013; Hubert et al. 2015, 2019; Lim et al. 2016; Beck et al.
632 2017; Dahrudin et al. 2017; Hutama et al. 2017; Nurul Farhana et al. 2018). The
633 establishment of taxa from mainland Asia during the Miocene and Pliocene supports the pre-
634 Pleistocene colonization hypothesis (Dodson et al. 1995; de Bruyn et al. 2013; Hendriks et al.
635 2019). These results agree that regional shallow seas only started to dry out during the
636 Pliocene, offering the possibility for dispersal of freshwater ancestors from mainland Asia.
637 During these periods, Borneo was certainly more connected to mainland Southeast Asia than
638 Sumatra and Java (Figs. 1B & 1C), and Borneo probably played a role in the initial diversity
639 build-up of Sundaland biotas (de Bruyn et al. 2013, 2014). From our inferences, Sumatra and
640 Java were first colonized from Borneo around 2.96 Ma via the North Sunda river system and
641 around 1.18 Ma via the East Sunda river system, respectively. Thus, our study reinforces the
642 conclusion of De Bruyn et al. (2014) that Borneo was the origin of most insular Sundaland
643 freshwater fish diversity.

644 Although the fish lineages we studied were present in Sundaland before the
645 Pleistocene, most of the MOTUs delimited here result from Pleistocene diversification events
646 that are expected to be affected by PCF (Esselstyn et al. 2009). Looking at the geography of
647 speciation for each group, we estimated that speciation of *Clarias* and *Channa* involved more
648 dispersal events compared to *Glyptothorax* and *Zenarchopteridae*, probably due to higher

649 dispersal ability (Kottelat et al. 1993; Berra 2001; Downing Meisner 2001; Pouyaud et al.
650 2009; Adamson et al. 2010; Jiang et al. 2011; Tan and Lim 2013; Serrao et al. 2014; Lim et
651 al. 2016; Ng and Kottelat 2016; Conte-Grand et al. 2017; Hutama et al. 2017; Nurul Farhana
652 et al. 2018). The dispersal patterns and absence of decline in the diversification rates suggest
653 that land bridges and the subsequent Pleistocene eustatic fluctuations drove further *in situ*
654 diversification with the interplay of the insular and paleoriver watersheds boundaries, as well
655 as habitat heterogeneity, which likely conditioned dispersal. Episodes of sea level rise during
656 interglacial periods might be expected to have fragmented Sundaland. Rising seas created
657 refugial insular areas in Borneo, Sumatra and Java. These saltwater barriers may have driven
658 further within-island diversification, explaining the numerous *in situ* diversification events
659 after long-distance dispersals in each of the four lineages. Yet, nearly half of Sundaland
660 Pleistocene speciation events involve dispersal between different islands. One possible
661 explanation is to consider the restored interconnectivity of both Sundaland land bridges and
662 paleoriver systems during Pleistocene glacial periods. Considering cooler and drier climate
663 during Pleistocene glaciations, it has been proposed that savanna and seasonal forest corridors
664 expanded through the interior of Sundaland, notably in East Sunda (Heaney 1992; Bird et al.
665 2005). Such inter-island bridges might be not easily penetrable by evergreen forest-dependent
666 taxa (Heaney 1992; Gorog et al. 2004; Bird et al. 2005; Pouyaud et al. 2009; Wurster et al.
667 2019). For example, despite its exemplary ability to disperse, *Clarias* was more likely to
668 speciate within an island (55.6%) than between islands (44.4%), which is supported by the
669 general phylogenetic division of the genus between black-water lineages and white-clear
670 water lineages. Black water lineages of *Clarias* are unlikely to switch habitats, inhibiting
671 them to easily penetrate freshwater habitats within the savanna/seasonal forest corridors
672 during glaciation periods (Pouyaud et al. 2009). Zenarchopteridae was the lineage least likely
673 to disperse between islands (18.3 %), most probably due to the poor dispersal ability of these

674 mostly ovoviviparous fishes (de Bruyn et al. 2013). Inter-island dispersal seems to be
675 correlated not only with dispersal ability, but also with habitat specificity (Heaney 2007).

676 The substantial proportion of founder-effect speciation between islands coupled with
677 dispersal between paleoriver systems and of speciation between paleorivers argues in favor of
678 the importance of habitat heterogeneity/diversity in the diversification of Sundaland
679 freshwater fishes. If paleoriver systems act as dispersal channels, within or between islands,
680 one can also argue that the borders of both paleorivers and insular systems might not be as
681 clearly delineated as previously supposed, due to biological aspects including the difference
682 in vegetation cover and/or physical aspects such as island geomorphology (Pouyaud et al.
683 2009; Brown et al. 2013; Hutama et al. 2017). The Bangka-Belitung islands are located at the
684 boundary between North Sunda and East Sunda river systems. We found no clear signal of its
685 affiliation with either of the two paleorivers. Similarly, borders between the two river systems
686 also could not be recovered strictly by biogeographic estimates in Lampung (eastern Sumatra)
687 and the southwestern part of Borneo, probably because the topography of these lowlands
688 facilitates gene flow between them. Notable examples for the indistinct boundaries between
689 the North Sunda and East Sunda paleorivers are exemplified by: (1) the existence of sister
690 lineages of *G. platypogonides* 1 and *G. stibaros* (Borneo), (2) dispersal during cladogenesis
691 between the *G. pictus* and *G. major* groups in Borneo, and (3) the trans-paleoriver
692 distributions of *Clarias leiacanthus* (Sumatra) and *C. meladerma* 1 (Sumatra and Borneo).
693 The physical island geomorphology could also contribute to re-arrangements of paleoriver
694 watersheds limits through headwater capture events among paleorivers through time. The
695 Bornean part of the North Sunda and East Sunda river systems share the same headwater area
696 in the center of the island.

697 Habitat fragmentation by mountainous terrain has been identified as a major
698 geomorphological driver of *in situ* diversification in Java resulting in abundant cryptic

699 diversity (Dahrudin et al. 2017; Hutama et al. 2017; Hubert et al. 2019). The existence of a
700 single paleoriver system may have driven ancestral lineages on the island to diversify further
701 within isolated pockets of riparian environment. This suggests that extant freshwater fish
702 diversity in Java resulted from high rampant *in situ* diversification rather than immigration
703 from elsewhere (Nguyen et al. 2008; Pouyaud et al. 2009; Hubert et al. 2015, 2019; Kusuma
704 et al. 2016; Dahrudin et al. 2017; Hutama et al. 2017). *In situ* diversification within islands
705 are found in all lineages studied here (e.g. *Glyptothorax platypogon*, *G. robustus*,
706 *Dermogenys pusilla*, and *Channa gachua*; Fig. 4, Fig. S2, Table S2).

707

708 ***Robustness of the Inferences, Limits and Perspectives***

709 For all four taxa under study, we found that tree topologies are congruent with
710 previously published phylogenetic hypotheses (Pouyaud et al. 2009; Jiang et al. 2011; de
711 Bruyn et al. 2013; Conte-Grand et al. 2017). Both ML and Bayesian analyses retrieved similar
712 and robust topologies, and each genus was found to be monophyletic, including the three
713 genera of Zenarchopteridae, with *Nomorhamphus* being the sister group of *Dermogenys* and
714 *Nomorhamphus* + *Dermogenys* as the sister group of *Hemirhamphodon*, supporting Meisner
715 (2001) and Lovejoy et al. (2004). Minor differences concern the monophyly of *Dermogenys*,
716 well supported now, contrary to previous studies (de Bruyn et al. 2013), and likely resulting
717 from the incorporation of multiple outgroups here. For *Clarias*, our analysis supports the
718 monophyly of Asian *Clarias* as previously suggested (Pouyaud et al. 2009). By contrast, ML
719 phylogenetic reconstruction (Fig. 2A, Fig. S1) and the Bayesian species tree (Fig. 4A, Fig.
720 S2) support the reciprocal monophyly of black water species (Clade II) and non black-water
721 species (Clade III). Only *Clarias microstomus* Ng 2001 and *C. planiceps* Ng 1999 departed
722 from this general trend. For *Glyptothorax*, differences with the reference phylogeny (Jiang et
723 al. 2011) are only due to a recent revision of Sundaland *Glyptothorax* species that led to the

724 revision and description of new taxa (Ng and Kottelat 2016), that have been incorporated here
725 as well as the addition of supplementary sequences (Hutama et al. 2017). Our findings are
726 consistent with the occurrence of two distinct clades in Sundaland (Fig. 2B, Fig. 4B, Fig. S1-
727 S2) as suggested earlier (Jiang et al. 2011). Finally, the species tree of *Channa* recovered here
728 is consistent with the eight species groups previously recognized (Rüber et al. 2019). Our
729 results differ from previous studies in placing the *C. punctata* group and *C. gachua* group in
730 early-diverging position instead of the *C. micropeltes* group and *C. lucius* group (Conte-
731 Grand et al. 2017; Rüber et al. 2019).

732 At the MOTUs level, several nominal species are not monophyletic, which is probably
733 due to practical taxonomic limitations in which different lineages with no apparent
734 morphological differences might have been lumped together into the same nominal species as
735 in the polyphyletic *Clarias nieuhofii* (as previously observed by Pouyaud et al. 2009),
736 *Glyptothorax platypogonides* and *Channa gachua*, as well as the paraphyletic *Nomorhamphus*
737 *megarrhamphus* or several *Glyptothorax* species (Jiang et al. 2011; Ng and Kottelat 2016).
738 These results are consistent with recent findings about the existence of high levels of cryptic
739 diversity among Sundaland freshwater fishes (Nguyen et al. 2008; Pouyaud et al. 2009;
740 Hubert et al. 2015, 2019; Conte-Grand et al. 2017; Dahruddin et al. 2017; Hutama et al. 2017;
741 Sholihah et al. 2020) which further points to the necessity of using genetic delimitation
742 methods for subsequent macroevolutionary analyses of complex biotas (Esselstyn et al. 2009;
743 Patel et al. 2011; Ruane et al. 2014; Hubert et al. 2015, 2019; Hutama et al. 2017). The
744 remarkable number of MOTUs recovered from nominal species (252 MOTUs vs. 110
745 morphological species) calls for more detailed taxonomic works on these taxa (Pouyaud et al.
746 2009; Hutama et al. 2017).

747 The divergence times we estimated using a molecular clock approach agree with
748 (*Hemirhamphodon*, *Dermogenys*, *Nomorhamphus*) or are younger than (*Clarias* and *Channa*)

749 previous estimates based on geological and fossil calibrations. We also provided the first
750 time-calibrated phylogeny for *Glyptothorax*. We found similar divergence times for
751 *Hemirhamphodon* (9.88 Ma, 95% HPD = 6.02-14.70 Ma, vs. 11.2 Ma, 95% HPD = 6.7-16
752 Ma; de Bruyn et al. 2013) and for the *Dermogenys* + *Nomorhamphus* clade (6.28 Ma, 95%
753 HPD = 4.28-8.66 Ma, vs. 13.3 Ma, 95% HPD = 8-18 Ma; de Bruyn et al. 2013). Younger
754 divergence estimates are found in *Clarias* and *Channa*: we estimated the Asian *Clarias* clade
755 to be 6.37 Ma (95% HPD = 3.99-9.19 Ma) while Pouyaud et al. (2009) dated the divergence
756 at 33.4 Ma. In the same way, we found a younger divergence time for *Channa* (14.35 Ma,
757 95% HPD = 10.54-18.41 Ma) than a previous estimate (28 Ma, 95% HPD = 24-32 Ma; Rüber
758 et al. 2020). A potential cause of these discrepancies is likely our use of a mixed
759 coalescent/diversification model here for the Bayesian species trees reconstruction that
760 accounts for the heterogeneity of absolute substitution rates within and between species (Ho
761 and Larson 2006). Within species, clock rates can be considerably higher than between
762 species because observed substitution rates are close to mutation rates, genetic drift being
763 later responsible for the loss of haplotypes leading to stationary substitution rates that are
764 most commonly used in phylogenetic reconstructions and molecular age estimates (Ho and
765 Larson 2006). Along the same lines, the use of alternative clock rates moderately impacted
766 our main results: (1) most MOTUs originated during the Pleistocene regardless of the clock
767 rate used; (2) diversification model selection was consistent across a wide range of clock
768 rates. Furthermore, this study is based on genetic recognition of MOTUs instead of
769 morphological species; the high proportion of cryptic diversity found here was not properly
770 accounted for by the previous phylogenetic reconstructions (Pouyaud et al. 2009; de Bruyn et
771 al. 2013). It can be expected that such a difference in taxon sampling has impacted absolute
772 age estimates. The general concordance in the distribution through time of MOTUs age

773 among taxa with different life history traits and origins argue in favor of the robustness of the
774 age estimates established here.

775 Finally, several issues can be highlighted that warrant further studies. First, although
776 the samples generally cover a broad range of localities around Sundaland, we identify
777 sampling gaps in large areas, including Sumatra for all taxa and in Peninsular Malaysia for
778 *Clarias*. Second, lack of COI sequences deposited in BOLD limited the application of RESL
779 algorithm and designation of Barcode Index Numbers (BINs) for older datasets (*Clarias* and
780 *Channa*). While there was no significant challenge from missing BINs during species
781 delimitation, more complete BIN would increase confidence in the species delimitation
782 results. The subfamily Rasborinae (Brittan 1972; Sholihah et al. 2020) could be an additional
783 biological model to assess the evolutionary history of Sundaland freshwater fishes. It has
784 multiple species-rich genera of small size (Liao et al. 2011; Tan and Armbruster 2018).
785 Lastly, the constant diversification rates through time inferred for two genera might result
786 from inadequate sampling. It has also been suggested that the DDD model might be biased
787 toward spurious detection of early explosive speciation dynamics resulting from the
788 underestimation of the actual number of evolutionary lineages created by taxonomic and
789 sampling bias (Barraclough and Nee 2001; Rabosky and Lovette 2008). Alternatively,
790 constant diversification rates have been a recurrent findings in tropical biomes (Esselstyn et
791 al. 2009; Patel et al. 2011). In this case, we have emphasized our effort to avoid clade-specific
792 taxonomic hypotheses (Esselstyn et al. 2009) by utilizing species delimitation methods for all
793 clades. We have detected numerous cryptic lineages in all taxa which might explain the
794 constant diversification rate found here, as proposed by Patel et al. (2011) to explain the
795 constant diversification rates of Neotropical *Pteroglossus*.

796

797 **CONCLUSIONS**

798 Our results indicate that the diversification and biogeography of freshwater fishes on
799 Sundaland are not solely dependent on Pleistocene sea-level fluctuations and associated
800 paleoriver systems, but are also affected by: (1) pre-Pleistocene geological history of
801 Sundaland, both at the origin of the development of paleoriver systems and responsible for
802 ancient speciation events among paleorivers; (2) the idiosyncratic effects of island boundaries
803 and the extent of paleorivers on diversification in each system; (3) ecology of the
804 paleoenvironments (Heaney 2007) and the (re)emerged paleorivers that were running through
805 the land bridges of Sundaland through the Pleistocene (Heaney 1992; Bird et al. 2005; Slik et
806 al. 2011); (4) the geomorphology of emerged and currently submerged Sundaland land
807 masses, creating indistinct borders among paleoriver systems; and (5) different evolutionary
808 responses of different groups with specific life history traits. Furthermore, it has been
809 suggested recently that the subsidence of Sundaland may have triggered the merge of
810 Sundaland landmasses during glacial times only 400,000 years ago, implying that eustatic
811 fluctuations prior to that period only marginally impacted the extent of emerged lands
812 (Husson et al. 2020). Although, not in line with the initial framework used here for testing the
813 Paleoriver Hypothesis, our observations concerning the significant effect of pre-Pleistocene
814 geological settlement and paleorivers arrangement on biodiversity are in agreement with this
815 recent finding. Finally, our findings propose new perspectives on the biogeography of
816 Sundaland freshwater biotas and open new questions about the interplay between geology and
817 paleoecology during dispersal and colonization of Sundaland islands. Riverine organisms
818 have constrained dispersal routes, whereas terrestrial organisms do not. As such, our studies
819 warrant new researches on the biogeography of terrestrial rainforest biotas of Sundaland.

820

821 **ACKNOWLEDGEMENTS**

822 This paper is part of the doctoral study of AS, funded by a full scholarship of the
823 Indonesian Endowment Fund for Education (LPDP) from the Indonesian Ministry of Finance.
824 The authors thank Hari Sutrisno and Cahyo Rahmadi as well as the staff of the Zoology
825 division of the Research Centre for Biology (Indonesian Institute of Sciences) for hosting AS
826 during the first year of her PhD. We thank David J. Lohman and one anonymous reviewer for
827 the relevant comments on earlier versions of the manuscript. This publication is ISEM
828 publication number 2020-230 SUD.

829

830 **References**

- 831 Adamson E.A.S., Hurwood D.A., Mather P.B. 2010. A reappraisal of the evolution of Asian
832 snakehead fishes (Pisces, Channidae) using molecular data from multiple genes and
833 fossil calibration. *Mol. Phylogenet. Evol.* 56:707–717.
- 834 Alonso D., Etienne R.S., McKane A.J. 2006. The merits of neutral theory. *Trends Ecol. Evol.*
835 21:451–457.
- 836 Alter S.E., Munshi-South J., Stiassny M.L.J. 2017. Genomewide SNP data reveal cryptic
837 phylogeographic structure and microallopatric divergence in a rapids-adapted clade of
838 cichlids from the Congo River. *Mol. Ecol.* 26:1401–1419.
- 839 April J., Hanner R.H., Dion-Côté A.M., Bernatchez L. 2012. Glacial cycles as an allopatric
840 speciation pump in North eastern american freshwater fishes. *Mol. Ecol.*
- 841 Barluenga M., Meyer A. 2004. The midas cichlid species complex: incipient sympatric
842 speciation in Nicaraguan cichlid fishes ? *Mol. Ecol.* 13:2061–2076.
- 843 Barraclough T.G., Nee S. 2001. Phylogenetics and Speciation. *Trends Ecol. Evol.* 16:391–
844 399.
- 845 Bates M., Hackett, S., J., Cracraft, J. J. 1998. Area-relationships in the neotropical lowlands:
846 an hypothesis based on raw distributions of passerine birds. *J. Biogeogr.* 25:783–793.

847 Beck S. V., Carvalho G.R., Barlow A., Ruber L., Hui Tan H., Nugroho E., Wowor D., Mohd
848 Nor S.A., Herder F., Muchlisin Z.A., de Bruyn M. 2017. Plio-Pleistocene
849 phylogeography of the Southeast Asian Blue Panchax killifish, *Aplocheilichthys panchax*.
850 PLoS One. 12:e0179557.

851 Berra T.M. 2001. Freshwater Fish Distribution. San Diego: Academic Press.

852 Betancur-R R., Wiley E.O., Arratia G., Acero A., Bailly N., Miya M., Lecointre G., Ortí G.
853 2017. Phylogenetic classification of bony fishes. BMC Evol. Biol. 17:162.

854 Bird M.I., Taylor D., Hunt C. 2005. Palaeoenvironments of insular Southeast Asia during the
855 Last Pleistocene. Glacial Period: a savanna corridor in Sundaland? . Quat. Sci. Rev.
856 24:2228–2242.

857 Blaxter M., Mann J., Chapman T., Thomas F., Whitton C., Floyd R., Abebe E. 2005. Defining
858 operational taxonomic units using DNA barcode data. Philos. Trans. R. Soc. B Biol. Sci.
859 360:1935–1943.

860 Bouckaert R., Heled J., Kühnert D., Vaughan T., Wu C.H., Xie D., Suchard M.A., Rambaut
861 A., Drummond A.J. 2014. BEAST 2: A Software Platform for Bayesian Evolutionary
862 Analysis. PLoS Comput. Biol. 10:1–6.

863 Brittan M.R. 1972. *Rasbora*: A revision of the Indo-Malayan freshwater fish genus Rasbora.
864 Hongkong: TFH Publications.

865 Brown R.M., Siler C.D., Oliveros C.H., Esselstyn J.A., Diesmos A.C., Hosner P.A., Linkem
866 C.W., Barley A.J., Oaks J.R., Sanguila M.B., Welton L.J., Blackburn D.C., Moyle R.G.,
867 Townsend Peterson A., Alcalá A.C. 2013. Evolutionary Processes of Diversification in a
868 Model Island Archipelago. Annu. Rev. Ecol. Evol. Syst. 44:411–435.

869 de Bruyn M., Rüber L., Nyländer S., Stelbrink B., Lovejoy N.R., Lavoué S., Heok Hui T.,
870 Nugroho E., Wowor D., Ng P.K.L., Siti Azizah M.N., Von Rintelen T., Hall R.,
871 Carvalho G.R. 2013. Paleo-drainage basin connectivity predicts evolutionary

872 relationships across three Southeast Asian biodiversity hotspots. *Syst. Biol.* 62:398–410.

873 de Bruyn M., Stelbrink B., Morley R.J., Hall R., Carvalho G.R., Cannon C.H., van den Bergh
874 G., Meijaard E., Metcalfe I., Boitani L., Maiorano L., Shoup R., von Rintelen T. 2014.
875 Borneo and Indochina are Major Evolutionary Hotspots for Southeast Asian
876 Biodiversity. *Syst Biol.* 63:23.

877 De Bruyn M., Wilson J.A., Mather P.B. 2004. Huxley's line demarcates extensive genetic
878 divergence between eastern and western forms of the giant freshwater prawn,
879 *Macrobrachium rosenbergii*. *Mol. Phylogenet. Evol.* 30:251–257.

880 Burbrink F.T., Chan Y.L., Myers E.A., Ruane S., Smith B.T., Hickerson M.J. 2016.
881 Asynchronous demographic responses to Pleistocene climate change in Eastern Nearctic
882 vertebrates. *Ecol. Lett.* 19:1457–1467.

883 Cannon C.H., Morley R.J., Bush M.B. 2009. The current refugial rainforests of Sundaland are
884 unrepresentative of their biogeographic past and highly vulnerable to disturbance. *Proc.*
885 *Natl. Acad. Sci. USA.* 106:11188–11193.

886 Clark J.R., Ree R.H., Alfaro M.E., King M.G., Wagner W.L., Roalson E.H. 2008. A
887 comparative study in ancestral range reconstruction methods: retracing the uncertain
888 histories of insular lineages. *Syst. Biol.* 57:693–707.

889 Clements R., Sodhi N.S., Schilthuizen M., Ng P.K.L. 2006. Limestone karsts of Southeast
890 Asia: imperiled arks of biodiversity. *Bioscience.* 56:733–742.

891 Condamine F.L., Nagalingum N.S., Marshall C.R., Morlon H. 2015a. Origin and
892 diversification of living cycads: a cautionary tale on the impact of the branching process
893 prior in Bayesian molecular dating. *BMC Evol. Biol.* 15:65.

894 Condamine F.L., Rolland J., Morlon H. 2013. Macroevolutionary perspectives to
895 environmental change. *Ecol. Lett.* 16:72–85.

896 Condamine F.L., Rolland J., Morlon H. 2019. Assessing the causes of diversification

897 slowdowns: temperature-dependent and diversity-dependent models receive equivalent
898 support. *Ecol. Lett.* 22:1900–1912.

899 Condamine F.L., Toussaint E.F.A., Clamens A.L., Genson G., Sperling F.A.H., Kergoat G.J.
900 2015b. Deciphering the evolution of birdwing butterflies 150 years after Alfred Russel
901 Wallace. *Sci. Rep.* 5:1–11.

902 Conte-Grand C., Britz R., Dahanukar N., Raghavan R., Pethiyagoda R., Tan H.H., Hadiaty
903 R.K., Yaakob N.S., Rüber L. 2017. Barcoding snakeheads (Teleostei, Channidae)
904 revisited: Discovering greater species diversity and resolving perpetuated taxonomic
905 confusions. *PLoS One.* 12:e0184017.

906 Currie D.J., Mittelbach G.G., Cornell H. V., Field R., Guegan J.-F., Hawkins B.A., Kaufman
907 D.M., Kerr J.T., Oberdorff T., O’Brien E., Turner J.R.G. 2004. Predictions and tests of
908 climate-based hypotheses of broad-scale variation in taxonomic richness. *Ecol. Lett.*
909 7:1121–1134.

910 Dahrudin H., Hutama A., Busson F., Sauri S., Hanner R., Keith P., Hadiaty R., Hubert N.
911 2017. Revisiting the ichthyodiversity of Java and Bali through DNA barcodes:
912 taxonomic coverage, identification accuracy, cryptic diversity and identification of
913 exotic species. *Mol. Ecol. Resour.* 17:288–299.

914 Van Dam M.H., Matzke N.J. 2016. Evaluating the influence of connectivity and distance on
915 biogeographical patterns in the south-western deserts of North America. *J. Biogeogr.*
916 43:1514–1532.

917 Darriba D., Taboada G.L., Doallo R., Posada D. 2015. Europe PMC Funders Group
918 jModelTest 2 : more models , new heuristics and high- performance computing. 9:6–9.

919 Dodson J.J., Colombani F., Ng P.K.L. 1995. Phylogeographic structure in mitochondrial
920 DNA of a South-east Asian freshwater fish, *Hemibagrus nemurus* (Sluroidei; Bagridae)
921 and Pleistocene sea-level changes on the Sunda shelf. *Mol. Ecol.* 4:331–346.

922 Downing Meisner A. 2001. Phylogenetic systematics of the viviparous halfbeak genera
923 Dermogenys and Nomorhamphus (Teleostei: Hemiramphidae: Zenarchopterinae). Zool.
924 J. Linn. Soc. 133:199–283.

925 Duchêne S., Lanfear R., Ho S.Y.W. 2014. The impact of calibration and clock-model choice
926 on molecular estimates of divergence times. Mol. Phylogenet. Evol. 78:277–289.

927 Edgar R.C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high
928 throughput. Nucleic Acids Res. 32:1792–7.

929 Esselstyn J.A., Brown R.M. 2009. The role of repeated sea-level fluctuations in the generation
930 of shrew (Soricidae: *Crocidura*) diversity in the Philippine Archipelago. Mol.
931 Phylogenet. Evol. 53:171–181.

932 Esselstyn J.A., Timm R.M., Brown R.M. 2009. Do geological or climatic processes drive
933 speciation in dynamic archipelagos? the tempo and mode of diversification in southeast
934 asian shrews. Evolution. 63:2595–2610.

935 Etienne R.S., Haegeman B., Stadler T., Aze T., Pearson P.N., Purvis A., Phillimore A.B.
936 2012. Diversity-dependence brings molecular phylogenies closer to agreement with the
937 fossil record. Proc. R. Soc. B Biol. Sci. 279:1300–1309.

938 Fujisawa T., Barraclough T.G. 2013. Delimiting species using single-locus data and the
939 generalized mixed yule coalescent approach: A revised method and evaluation on
940 simulated data sets. Syst. Biol. 62:707–724.

941 Gathorne-Hardy F.J., Syaukani, Davies R.G., Eggleton P., Jones D.T. 2002. Quaternary
942 rainforest refugia in south-east Asia: using termites (Isoptera) as indicators. Biol. J. Linn.
943 Soc. 75:453–466.

944 Gavrilets S., Vose A. 2006. Dynamic patterns of adaptative radiation. Proc. Natl. Acad. Sci.
945 USA. 102:18040–18045.

946 Giam X., Koh L.P., Tan H.H., Miettinen J., Tan H.T.W., Ng P.K.L. 2012. Global extinctions

947 of freshwater fishes follow peatland conversion in Sundaland. *Front. Ecol. Environ.*
948 10:465–470.

949 Gorog A.J., Sinaga M.H., Engstrom M.D. 2004. Vicariance or dispersal? Historical
950 biogeography of three Sunda shelf marine rodents (*Maxomys surifer*, *Leopoldamys*
951 *sabanus* and *Maxomys whiteheadi*). *Biol. J. Linn. Soc.* 81:91–109.

952 Guindon S., Gascuel O. 2003. A Simple, Fast, and Accurate Algorithm to Estimate Large
953 Phylogenies by Maximum Likelihood. *Syst. Biol.* 52:696–704.

954 Guyot J.L. 1993. Hydrogeochemie des fleuves de l'Amazonie Bolivienne. .

955 Haffer J. 1997. Alternative models of vertebrate speciation in Amazonia: an overview.
956 *Biodivers. Conserv.* 6:26.

957 Hall R. 2009. Southeast Asia's changing palaeogeography. *Blumea J. Plant Taxon. Plant*
958 *Geogr.* 54:148–161.

959 Hall R. 2013. The palaeogeography of Sundaland and Wallacea since the Late Jurassic. *J.*
960 *Limnol.* 72:1–17.

961 Hardman M., Lundberg J.G. 2006. Molecular phylogeny and chronology of diversification
962 for “phractocephaline” catfishes (Siluriformes: Pimelodidae) based on mitochondrial
963 DNA and DNA recombination activating gene 2 sequences. *Mol. Phylogenet. Evol.*
964 40:410–418.

965 Heaney L.R. 1992. A synopsis of climatic and vegetational change in Southeast Asia. *Trop.*
966 *For. Clim.:*53–61.

967 Heaney L.R. 2007. Is a new paradigm emerging for oceanic island biogeography? *J.*
968 *Biogeogr.* 34:753–757.

969 Heled J., Drummond A.J. 2010. Bayesian Inference of Species Trees from Multilocus Data.
970 *Mol. Biol. Evol.* 27:570–580.

971 Hendriks K.P., Alciatore G., Schilthuizen M., Etienne R.S. 2019. Phylogeography of Bornean

972 land snails suggests long-distance dispersal as a cause of endemism. *J. Biogeogr.*

973 Herder F., Nolte A.W., Pfaender J., Schwarzer J., Hadiaty R.K., Schliwen U.K. 2006.

974 Adaptative radiation and hybridization in Wallace's dreamponds: evidence from sailfin
975 silversides in the Malili Lakes of Sulawesi. *Proc. R. Soc. B.* 273:2209–2217.

976 Ho S.Y.W., Larson G. 2006. Molecular clocks: when times are a-changin'. *TRENDS Genet.*
977 22:79–83.

978 Hoffmann M., Hilton-Taylor C., Angulo A., Böhm M., Brooks T.M., Butchart S.H.M.,
979 Carpenter K.E., Chanson J., Collen B., Cox N.A., Darwall W.R.T., Dulvy N.K., Harrison
980 L.R., Katariya V., Pollock C.M., Quader S., Richman N.I., Rodrigues A.S.L., Tognelli
981 M.F., Vié J.C., Aguiar J.M., Allen D.J., Allen G.R., Amori G., Ananjeva N.B., Andreone
982 F., Andrew P., Ortiz A.L.A., Baillie J.E.M., Baldi R., Bell B.D., Biju S.D., Bird J.P.,
983 Black-Decima P., Blanc J.J., Bolaños F., Bolivar-G. W., Burfield I.J., Burton J.A.,
984 Capper D.R., Castro F., Catullo G., Cavanagh R.D., Channing A., Chao N.L., Chenery
985 A.M., Chiozza F., Clausnitzer V., Collar N.J., Collett L.C., Collette B.B., Cortez
986 Fernandez C.F., Craig M.T., Crosby M.J., Cumberlidge N., Cuttelod A., Derocher A.E.,
987 Diesmos A.C., Donaldson J.S., Duckworth J.W., Dutton G., Dutta S.K., Emslie R.H.,
988 Farjon A., Fowler S., Freyhof J., Garshelis D.L., Gerlach J., Gower D.J., Grant T.D.,
989 Hammerson G.A., Harris R.B., Heaney L.R., Hedges S.B., Hero J.M., Hughes B.,
990 Hussain S.A., Icochea M. J., Inger R.F., Ishii N., Iskandar D.T., Jenkins R.K.B., Kaneko
991 Y., Kottelat M., Kovacs K.M., Kuzmin S.L., La Marca E., Lamoreux J.F., Lau M.W.N.,
992 Lavilla E.O., Leus K., Lewison R.L., Lichtenstein G., Livingstone S.R., Lukoschek V.,
993 Mallon D.P., McGowan P.J.K., McIvor A., Moehlman P.D., Molur S., Alonso A.M.,
994 Musick J.A., Nowell K., Nussbaum R.A., Olech W., Orlov N.L., Papenfuss T.J., Parra-
995 Olea G., Perrin W.F., Polidoro B.A., Pourkazemi M., Racey P.A., Ragle J.S., Ram M.,
996 Rathbun G., Reynolds R.P., Rhodin A.G.J., Richards S.J., Rodríguez L.O., Ron S.R.,

997 Rondinini C., Rylands A.B., de Mitcheson Y.S., Sanciangco J.C., Sanders K.L., Santos-
 998 Barrera G., Schipper J., Self-Sullivan C., Shi Y., Shoemaker A., Short F.T., Sillero-
 999 Zubiri C., Silvano D.L., Smith K.G., Smith A.T., Snoeks J., Stattersfield A.J., Symes
 1000 A.J., Taber A.B., Talukdar B.K., Temple H.J., Timmins R., Tobias J.A., Tsytsulina K.,
 1001 Tweddle D., Ubeda C., Valenti S. V., Van Dijk P.P., Veiga L.M., Veloso A., Wege D.C.,
 1002 Wilkinson M., Williamson E.A., Xie F., Young B.E., Akçakaya H.R., Bennun L.,
 1003 Blackburn T.M., Boitani L., Dublin H.T., da Fonseca G.A.B., Gascon C., Lacher T.E.,
 1004 Mace G.M., Mainka S.A., McNeely J.A., Mittermeier R.A., Reid G.M.G., Rodriguez
 1005 J.P., Rosenberg A.A., Samways M.J., Smart J., Stein B.A., Stuart S.N. 2010. The impact
 1006 of conservation on the status of the world's vertebrates. *Science* (80-.). 330:1503–1509.
 1007 Hubbell S.P. 2001. *The Unified Neutral Theory of Biodiversity and Biogeography*. Princeton:
 1008 Princeton University Press.
 1009 Hubert N., Calcagno V., Etienne R.S., Mouquet N. 2015. Metacommunity speciation models
 1010 and their implications for diversification theory. *Ecol. Lett.* 18:864–881.
 1011 Hubert N., Lumbantobing D., Sholihah A., Dahruddin H., Delrieu-Trottin E., Busson F., Sauri
 1012 S., Hadiaty R., Keith P. 2019. Revisiting species boundaries and distribution ranges of
 1013 *Nemacheilus* spp. (Cypriniformes: Nemacheilidae) and *Rasbora* spp. (Cypriniformes:
 1014 Cyprinidae) in Java, Bali and Lombok through DNA barcodes: implications for
 1015 conservation in a biodiversity hotspot. *Conserv. Genet.* 20:517–529.
 1016 Hubert N., Renno J.F. 2006. Historical Biogeography of South American Freshwater fishes. *J.*
 1017 *Biogeogr.* 33:1414–1436.
 1018 Husson L., Boucher F.C., Sarr A.C., Sepulchre P., Cahyarini S.Y. 2020. Evidence of
 1019 Sundaland's subsidence requires revisiting its biogeography. *J. Biogeogr.* 47:843–853.
 1020 Hutama A., Dahruddin H., Busson F., Sauri S., Keith P., Hadiaty R.K., Hanner R., Suryobroto
 1021 B., Hubert N. 2017. Identifying spatially concordant evolutionary significant units across

1022 multiple species through DNA barcodes: Application to the conservation genetics of the
1023 freshwater fishes of Java and Bali. *Glob. Ecol. Conserv.* 12:170–187.

1024 Jiang W., Ng H.H., Yang J., Chen X. 2011. Monophyly and phylogenetic relationships of the
1025 catfish genus *Glyptothorax* (Teleostei: Sisoridae) inferred from nuclear and
1026 mitochondrial gene sequences. *Mol. Phylogenet. Evol.* 61:278–289.

1027 Kadarusman, Hubert N., Hadiaty R.K., Sudarto, Paradis E., Pouyaud L. 2012. Cryptic
1028 diversity in indo-australian rainbowfishes revealed by DNA barcoding: Implications for
1029 conservation in a biodiversity hotspot candidate. *PLoS One.* 7.

1030 Kekkonen M., Hebert P.D.N. 2014. DNA barcode-based delineation of putative species:
1031 efficient start for taxonomic workflows. *Mol. Ecol. Resour.* 14:706–715.

1032 Kekkonen M., Mutanen M., Kaila L., Nieminen M., Hebert P.D.N. 2015. Delineating Species
1033 with DNA Barcodes: A Case of Taxon Dependent Method Performance in Moths. *PLoS*
1034 *One.* 10:e0122481.

1035 Kisel Y., McInnes L., Toomey N.H., Orme C.D.L. 2011. How diversification rates and
1036 diversity limits combine to create large-scale species-area relationships. *Philos. Trans. R.*
1037 *Soc. B Biol. Sci.* 366:2514–2525.

1038 Knowlton N., Weight L.A., Solorzano L.A., Mills D.K., Bermingham E. 1992. Divergence of
1039 proteins, mitochondrial DNA and reproductive compatibility across the Isthmus of
1040 Panama. *Science* (80-.). 260:1629–1632.

1041 Kottelat M., Whitten A.J., Kartikasari N., Wirjoatmodjo S. 1993. *Freshwater Fishes of*
1042 *Western Indonesia and Sulawesi.* Periplus Edition (HK) Ltd.

1043 Kumar S., Stecher G., Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis
1044 Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* 33:1870–4.

1045 Kusuma W.E., Ratmuangkhwang S., Kumazawa Y. 2016. Molecular phylogeny and historical
1046 biogeography of the Indonesian freshwater fish *Rasbora lateristriata* species complex

1047 (Actinopterygii: Cyprinidae): Cryptic species and west-to-east divergences. Mol.
1048 Phylogenet. Evol. 105:212–223.

1049 Landry L., Vincent W., Bernatchez L. 2007. Parallel evolution of lake whitefish dwarf
1050 ecotypes in association with limnological features of their adaptive landscape. J. Evol.
1051 Biol.

1052 Li F., Li S. 2018. Palaeocene-Eocene and Plio-Pleistocene sea-level changes as “species
1053 pumps” in Southeast Asia: Evidence from *Althepus* spiders. Mol. Phylogenet. Evol.
1054 127:545–555.

1055 Liao T.Y., Kullander S.O., Fang F. 2011. Phylogenetic position of rasborin cyprinids and
1056 monophyly of major lineages among the Danioninae, based on morphological characters
1057 (Cypriniformes: Cyprinidae). J. Zool. Syst. Evol. Res. 49:224–232.

1058 Lim H.-C., Abidin M.Z., Pulungan C.P., De Bruyn M., Mohd Nor S.A. 2016. DNA barcoding
1059 reveals high cryptic diversity of freshwater halfbeak genus *Hemirhamphodon* from
1060 Sundaland. PLoSONE. 11:e0163596.

1061 Lisiecki L.E., Raymo M.E. 2005. A Pliocene-Pleistocene stack of 57 globally distributed
1062 benthic δ 18O records. Paleoclimatology. 20:1–17.

1063 Liu J., Guo X., Chen D., Li J., Yue B., Zeng X. 2019. Diversification and historical
1064 demography of the rapid racerunner (*Eremias velox*) in relation to geological history and
1065 Pleistocene climatic oscillations in arid Central Asia. Mol Phylogenet Evol. 130:244–
1066 258.

1067 Lohman K., De Bruyn M., Page T., Von Rintelen K., Hall R., Ng P.K.L., Shih H.-T.,
1068 Carvalho G.R., Von Rintelen T. 2011. Biogeography of the Indo-Australian archipelago.
1069 Annu. Rev. Ecol. Evol. Syst. 42:205–226.

1070 Lovejoy N.R., Iranpour M., Collette B.B. 2004. Phylogeny and jaw ontogeny of beloniform
1071 fishes. Integr. Comp. Biol. 44:366–377.

1072 Maddison W.P., Maddison D.R. 2019. Mesquite: a modular system for evolutionary analysis.
1073 Version 3.61. <http://www.mesquiteproject.org>.

1074 Maina J.N., Maloiy G.M.O. 1986. The morphology of the respiratory organs of the African
1075 air-breathing catfish (*Clarias mossambicus*): A light, electron and scanning microscopic
1076 study, with morphometric observations. *J. Zool.* 209:421–445.

1077 Matzke N.J. 2013. Probabilistic historical biogeography: New models for founder-event
1078 speciation, imperfect detection, and fossils allow improved accuracy and model-testing.
1079 *Front. Biogeogr.* 5:242–248.

1080 Matzke N.J. 2014. Model selection in historical biogeography reveals that founder-event
1081 speciation is a crucial process in island clades. *Syst. Biol.* 63:951–970.

1082 McPeck M.A. 2008. The ecological dynamics of clade diversification and community
1083 assembly. *Am. Nat.* 172:e270–e284.

1084 Meisner A.D. 2001. Phylogenetic systematics of the viviparous halfbeak genera *Dermogenys*
1085 and *Nomorhamphus* (Teleostei: Hemiramphidae: Zenarchopterinae). *Zool. J. Linn. Soc.*
1086 133:199–283.

1087 Miller K.G., Kominz M.A., Browning J. V, Wright J.D., Mountain G.S., Katz M.E.,
1088 Sugarman P.J., Cramer B.S., Christie-Blick N., Pekar S.F. 2005. The phanerozoic record
1089 of global sea-level change. *Science* (80-.). 310:1293–1298.

1090 Miller M.A., Pfeiffer W., Schwartz T. 2010. Creating the CIPRES Science Gateway for
1091 inference of large phylogenetic trees. 2010 Gatew. Comput. Environ. Work. GCE 2010.

1092 Mittelbach G.G., Schemske D.W., Cornell H. V, Allen A.P., Brown J.M., Bush M.B.,
1093 Harrison S.P., Hurlbert A.H., Knowlton N., Lessios H.A., McCain C.M., McCune A.R.,
1094 McDade L.A., McPeck M.A., Near T.J., Price T.D., Ricklefs R.E., Roy K., Sax D.F.,
1095 Schluter D., Sobel J.M., Turelli M. 2007. Evolution and the latitudinal diversity gradient:
1096 speciation, extinction and biogeography. *Ecol. Lett.* 10:315–331.

- 1097 Morlon H., Lewitus E., Condamine F.L., Manceau M., Clavel J., Drury J. 2016. RPANDA: an
 1098 R package for macroevolutionary analyses on phylogenetic trees. *Methods Ecol. Evol.*
 1099 7:589–597.
- 1100 Morlon H., Parsons T.L., Plotkin J.B. 2011. Reconciling molecular phylogenies with the
 1101 fossil record. *Proc. Natl. Acad. Sci. U. S. A.* 108:16327–16332.
- 1102 Munshi J.S.D. 1961. The accessory respiratory organs of *Clarias batrachus* (Linn.). *J.*
 1103 *Morphol.* 109:115–139.
- 1104 Myers N., Mittermeier R.A., Mittermeier C.G., da Fonseca G.A.B., Kent F. 2000.
 1105 Biodiversity hotspots for conservation priorities. *Nature.* 403:853–858.
- 1106 Nagel L., Schluter D. 1998. Body size, natural selection, and speciation in sticklebacks.
 1107 *Evolution.* 52:209–218.
- 1108 Ng H.H., Kottelat M. 2016. The *Glyptothorax* of Sundaland: A revisionary study (Teleostei:
 1109 Sisoridae).
- 1110 Ng P.K.L., Tay J.B., Lim B.K. 1994. Diversity and conservation of blackwater fishes in
 1111 Peninsular Malaysia, particularly in the North Selangor peat swamp forest.
 1112 *Hydrobiologia.* 285:203–218.
- 1113 Nguyen T.T.T., Na-Nakorn U., Sukmanomon S., ZiMing C. 2008. A study on phylogeny and
 1114 biogeography of mahseer species (Pisces: Cyprinidae) using sequences of three
 1115 mitochondrial DNA gene regions. *Mol. Phylogenet. Evol.*
- 1116 Nolte A., Freyhof J., Stemshorn K., Tautz D. 2005. An invasive lineage of sculpins, *Cottus*
 1117 sp. (Pisces, Teleostei) in the Rhine with new habitat adaptations has originated from
 1118 hybridization between old phylogeographic groups. *Proc. R. Soc. London, B.* 272:2379–
 1119 2387.
- 1120 Nores M. 1999. An alternative hypothesis for the origin of Amazonian bird diversity. *J.*
 1121 *Biogeogr.* 26:475–485.

1122 Nores N. 2004. The implications of tertiary and quaternary sea level rise events for avian
1123 distribution patterns in the lowlands of northern South America. *Glob. Ecol. Biogeogr.*
1124 13:149–161.

1125 Nurul Farhana S., Muchlisin Z.A., Duong T.Y., Tanyaros S., Page L.M., Zhao Y., Adamson
1126 E.A.S., Khaironizam M.Z., de Bruyn M., Siti Azizah M.N. 2018. Exploring hidden
1127 diversity in Southeast Asia's *Dermogenys* spp. (Beloniformes: Zenarchopteridae)
1128 through DNA barcoding. *Sci. Rep.* 8:10787.

1129 Ogilvie H.A., Bouckaert R.R., Drummond A.J. 2017. StarBEAST2 brings faster species tree
1130 inference and accurate estimates of substitution rates. *Mol. Biol. Evol.* 34:2101–2114.

1131 Orti Meyer, A. G. 1997. The Radiation of Characiform Fishes and the limits of resolution of
1132 mitochondrial ribosomal DNA sequences. *Syst. Biol.* 46:75–100.

1133 Papadopoulou A., Knowles L.L. 2015a. Genomic tests of the species-pump hypothesis:
1134 Recent island connectivity cycles drive population divergence but not speciation in
1135 Caribbean crickets across the Virgin Islands. *Evolution.* 69:1501–1517.

1136 Papadopoulou A., Knowles L.L. 2015b. Species-specific responses to island connectivity
1137 cycles: Refined models for testing phylogeographic concordance across a Mediterranean
1138 Pleistocene Aggregate Island Complex. *Mol. Ecol.* 24:4252–4268.

1139 Paradis E. 2012. *Analysis of phylogenetics and evolution with R*. New York: Springer.

1140 Paradis E., Claude J., Strimmer K. 2004. APE: Analyses of phylogenetics and evolution in R
1141 language. *Bioinformatics.* 20:289–290.

1142 Paradis E., Schliep K. 2019. ape 5.0: an environment for modern phylogenetics and
1143 evolutionary analyses in R. *Bioinformatics.* 35:526–528.

1144 Patel S., Weckstein J.D., Patane J.S., Bates J.M., Aleixo A. 2011. Temporal and spatial
1145 diversification of *Pteroglossus aracaris* (AVES: Ramphastidae) in the neotropics:
1146 constant rate of diversification does not support an increase in radiation during the

- 1147 Pleistocene. *Mol Phylogenet Evol.* 58:105–115.
- 1148 Pellissier L., Leprieur F., Parravicini V., Cowman P.F., Kulbicki M., Litsios G., Olsen S.M.,
1149 Wisz M.S., Bellwood D.R., Mouillot D. 2014. Quaternary coral reef refugia preserved
1150 fish diversity. *Science* (80-). 344:1015–1019.
- 1151 Phillimore A.B., Price T.D. 2008. Density-dependent cladogenesis in birds. *PLoS Biol.* 6:e71.
- 1152 Pons J., Barraclough T.G., Gomez-Zurita J., Cardoso A., Duran D.P., Hazell S., Kamoun S.,
1153 Sumlin W.D., Vogler A.P. 2006. Sequence-based species delimitation for the DNA
1154 taxonomy of undescribed insects. *Syst. Biol.* 55:595–606.
- 1155 Pouyaud L., Sudarto, Paradis E. 2009. The phylogenetic structure of habitat shift and
1156 morphological convergence in Asian *Clarias* (Teleostei, Siluriformes: Clariidae). *J.*
1157 *Zool. Syst. Evol. Res.* 47:344–356.
- 1158 Puillandre N., Lambert A., Brouillet S., Achaz G. 2012. ABGD, Automatic Barcode Gap
1159 Discovery for primary species delimitation. *Mol. Ecol.* 21:1864–1877.
- 1160 Rabosky D.L., Lovette I.J. 2008. Explosive evolutionary radiations: Decreasing speciation or
1161 increasing extinction through time? *Evolution.* 62:1866–1875.
- 1162 Rambaut A., Drummond A.J., Xie D., Baele G., Suchard M.A. 2018. Posterior summarization
1163 in Bayesian phylogenetics using Tracer 1.7. *Syst. Biol.* 67:901–904.
- 1164 Ratnasingham S., Hebert P.D.N. 2007. BOLD: The Barcode of Life Data System
1165 (www.barcodinglife.org). *Mol. Ecol. Notes.* 7:355–364.
- 1166 Ratnasingham S., Hebert P.D.N. 2013. A DNA-Based Registry for All Animal Species: The
1167 Barcode Index Number (BIN) System. *PLoS One.* 8.
- 1168 Read C.I., Bellwood D.R., Herwerden van L. 2006. Ancient origins of Indo-Pacific coral reef
1169 fish biodiversity: a case study of the Leopard wrasses (Labridae: *Macropharyngodon*).
1170 *Mol. Phylogenet. Evol.* 38:808–819.
- 1171 Ree R.H., Sanmartín I. 2018. Conceptual and statistical problems with the DEC+ J model of

1172 founder-event speciation and its comparison with DEC via model selection. *J. Biogeogr.*
1173 45:741–749.

1174 Ruane S., Bryson R.W., Pyron R.A., Burbrink F.T. 2014. Coalescent species delimitation in
1175 Milksnakes (Genus *Lampropeltis*) and impacts on phylogenetic comparative analyses.
1176 *Syst. Biol.* 63:231–250.

1177 Rüber L., Tan H.H., Britz R. 2019. Snakehead (Teleostei: Channidae) diversity and the
1178 Eastern Himalaya biodiversity hotspot. *J. Zool. Syst. Evol. Res.*:1–31.

1179 Sathiamurthy E., Voris K.H. 2006. Maps of Holocene Sea Level Transgression and
1180 Submerged Lakes on the Sunda Shelf. .

1181 Schluter D. 2000. Ecological causes of adaptative radiation. *Am. Nat.* 148:40–64.

1182 Seehausen O. 2007. Evolution and ecological theory: Chance, historical contingency and
1183 ecological determinism jointly determine the rate of adaptive radiation. *Heredity (Edinb).*
1184 99:361–363.

1185 Seehausen O., Takimoto G., Roy D., Jokela J. 2008. Speciation reversal and biodiversity
1186 dynamics with hybridization in changing environments. *Mol. Ecol.* 17:30–44.

1187 Serrao N.R., Steinke D., Hanner R.H. 2014. Calibrating snakehead diversity with DNA
1188 barcodes: Expanding taxonomic coverage to enable identification of potential and
1189 established invasive species. *PLoS One.* 9.

1190 Sholihah A., Delrieu-Trottin E., Sukmono T., Dahruddin H., Risdawati R., Elvyra R.,
1191 Wibowo A., Kustiati K., Busson F., Sauri S., Nurhaman U., Dounias E., Zein M.S.A.,
1192 Fitriana Y., Utama I.V., Hubert N. 2020. Disentangling the taxonomy of the subfamily
1193 Rasborinae (Cypriniformes, Danionidae) in Sundaland using DNA barcodes. *Sci. Rep.*

1194 Slik J.W.F., Aiba S.-I., Bastian M., Brearley F.Q., Cannon C.H., Eichhorn K.A.O.,
1195 Fredriksson G., Kartawinata K., Laumonier Y., Mansor A., Marjokorpi A., Meijaard E.,
1196 Morley R.J., Nagamasu H., Nilus R., Nurtjahya E., Payne J., Permana A., Poulsen A.D.,

1197 Raes N., Riswan S., van Schaik C.P., Sheil D., Sidiyasa K., Suzuki E., van Valkenburg
 1198 J.L.C.H., Webb C.O., Wich S., Yoneda T., Zakaria R., Zweifel N. 2011. Soils on
 1199 exposed Sunda Shelf shaped biogeographic patterns in the equatorial forests of Southeast
 1200 Asia. *Proc. Natl. Acad. Sci. USA*. 108:12343–12347.

1201 Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of
 1202 large phylogenies. *Bioinformatics*. 30:1312–1313.

1203 Sullivan J., Lavoué S., Hopkins C. 2002. Discovery and phylogenetic analysis of e riverine
 1204 species flock of african electric fishes (Mormyridae: Teleostei). *Evolution*. 56:597–616.

1205 Tan H.H., Lim K.K.P. 2013. Three new species of freshwater halfbeaks (Teleostei:
 1206 Zenarchopteridae: *Hemirhamphodon*) from Borneo. *Raffles Bull. Zool*. 61:735–747.

1207 Tan M., Armbruster J.W. 2018. Phylogenetic classification of extant genera of fishes of the
 1208 order Cypriniformes (Teleostei: Ostariophysi). *Zootaxa*. 4476:6–39.

1209 Tan M.P., Jamsari A.F.J., Siti Azizah M.N. 2012. Phylogeographic Pattern of the Striped
 1210 Snakehead, *Channa striata* in Sundaland: Ancient River Connectivity, Geographical and
 1211 Anthropogenic Singnatures. *PLoS One*. 7:1–11.

1212 Verheyen E., Salzburger W., Snoeks J., Meyer A. 2003. Origin of the superflock of cichlid
 1213 fishes from Lake Victoria, East Africa. *Science* (80-.). 300:325–329.

1214 Voris H.K. 2000. Maps of Pleistocene sea levels in Southeast Asia: Shorelines, river systems
 1215 and time durations. *J. Biogeogr*. 27:1153–1167.

1216 Walker J.D., Geissman J.W., Bowring S.A., Babcock L.E. 2018. GSA Geologic Time Scale v.
 1217 5.0. .

1218 Weigelt P., Steinbauer M.J., Cabral J.S., Kreft H. 2016. Late quaternary climate change
 1219 shapes island biodiversity. *Nature*. 532:99–102.

1220 Wiens J.J., Donoghue M.J. 2004. Historical biogeography, ecology and species richness.
 1221 *Trends Ecol. Evol*. 19:639–644.

- 1222 Woodruff D.S. 2010. Biogeography and conservation in Southeast Asia: How 2.7 million
1223 years of repeated environmental fluctuations affect today's patterns and the future of the
1224 remaining refugial-phase biodiversity. *Biodivers. Conserv.* 19:919–941.
- 1225 Wurster C.M., Rifai H., Zhou B., Haig J., Bird M.I. 2019. Savanna in equatorial Borneo
1226 during the late Pleistocene. *Sci. Rep.* 9:1–7.
- 1227 Zhang J., Kapli P., Pavlidis P., Stamatakis A. 2013. A general species delimitation method
1228 with applications to phylogenetic placements. *Bioinformatics.* 29:2869–2876.

1229

1230 **FIGURE AND TABLE CAPTIONS**

1231 Figure 1. Paleogeographic maps of Sundaland in the last 20 million years Pleistocene sea
1232 levels and associated paleorivers. Reconstructions of historical land and sea distribution
1233 during the Neogene, depicting Sundaland during the early Miocene (A), late Miocene (B) and
1234 early Pliocene (C) (modified from Lohman et al. 2011). The Paleoriver systems in Southeast
1235 Asia are shown (D), depicting East Sunda (1), North Sunda (2), Malacca strait (3), Siam (4)
1236 river systems of Sundaland and the Mekong river system of IndoBurma. Lands exposed
1237 during -60 and -120m sea level decreases are illustrated in D. The cumulative number of
1238 sampling locations for each group within the paleoriver systems is provided (E) with color
1239 codes as follows: orange for *Clarias*, green for *Glyptothorax*, light blue for *Zenarchopteridae*
1240 and rose for *Channa*; and grey circles indicate that no sample was obtained from those areas.

1241

1242 Figure 2. Maximum likelihood trees and species delimitation for (A) *Clarias*, (B)
1243 *Glyptothorax*, (C) *Zenarchopteridae* and (D-E) *Channa*. For each clade, MOTUs delimited
1244 according to GMYC, PTP, BIN and ABGD (yellow), and the 50 percent consensus scheme of
1245 delimitation (red) are shown. Node circles represent bootstrap proportions and nodes

1246 numbering corresponds with the calibration nodes in Table 1. Major clades are labeled on the
1247 right.

1248

1249 Figure 3. Diversification through time of Sundaland freshwater fishes. Panel A shows the
1250 lineages through time (LTT) plot with confidence intervals collected from 1000 randomly
1251 sampled trees along the StarBEAST2 MCMC, superimposed with the total number of
1252 nominal species and MOTUs through time (0.5 Myr class). Panel B shows the speciation rate
1253 through time for each group, based on the most likely model of diversification (constant for
1254 *Clarias* and *Glyptothorax*, time-exponential for Zenarchopteridae, sea level-dependent for
1255 *Channa*). Panel C shows the sea level fluctuations over the last 15 Myr (adapted from Miller
1256 et al. 2005).

1257

1258 Figure 4. Panels a to d show the Bayesian maximum credibility trees for (A) *Clarias*, (B)
1259 *Glyptothorax*, (C) Zenarchopteridae and (D) *Channa* as well as the ancestral area
1260 reconstructions for each group, based on islands (left) and paleorivers (right). Panels E to F
1261 show exemplar specimens for (E) *Clarias*, (F) *Glyptothorax*, (G) Zenarchopteridae and (H)
1262 *Channa* with their relative size (scale bar = 1 cm).

1263

1264 Figure 5. Geographic pattern of speciation in Southeast Asian freshwater fishes. The plots
1265 show the numbers of speciation events associated to the four geographic patterns of speciation
1266 as follows: between islands and within paleorivers, between islands and between paleorivers,
1267 within islands and within paleorivers, and within islands and between paleorivers. Each bar
1268 shows the number of speciation events for each category for (A) all groups (cumulative), (B)
1269 Zenarchopteridae, (C) *Channa*, (D) *Clarias* and (E) *Glyptothorax*.

1270

1271 Table 1. Reference nodes and associated time calibrations for each group. Node numbering is
1272 as in Figure 2. Age estimates result from the Bayesian StarBEAST2 analyses and parameter
1273 distributions represent median age, 5 percent and 95 percent quantiles for each reference
1274 nodes

1275

1276 Table 2. Summary statistics of geographic patterns of speciation events for *Clarias*,
1277 *Glyptothorax*, Zenarchopteridae and *Channa* inferred using ancestral area estimation by
1278 BioGeoBEARS.

1279

1280 **SUPPLEMENTARY MATERIALS**

1281 Figure S1. Maximum Likelihood trees for *Clarias* (A), *Glyptothorax* (B), *Beloniformes* (C)
1282 and *Channa* (D) including tip labels as in Table S2 (online supplementary material)

1283

1284 Figure S2. StarBEAST2 species trees for *Clarias* (A), *Glyptothorax* (B), *Beloniformes* (C)
1285 and *Channa* (D) including tip labels as in Table S2 (online supplementary material)

1286

1287 Figure S3. Lineage-through time plots (A) and species age distributions (B) according to
1288 varying substitution rate hypotheses (clock rates ranging from 0.4 to 1.8 % divergence per
1289 million years) for *Clarias*, *Glyptothorax*, *Beloniformes* and *Channa*

1290

1291 Table S1. Description of the 17 diversification models tested for *Clarias*, *Glyptothorax*,
1292 *Beloniformes* and *Channa*, including acronym, rate variation, number of parameters and
1293 parameters involved (see material and methods section for details)

1294

1295 Table S2. List of the DNA sequences compiled in the present study including systematic
1296 (order, family, genus, species and updated species names), specimen number assigned,
1297 locality, geographic distribution (locality, island-based and palaeodrainage-based), GenBank
1298 accession numbers, DNA-based species delimitation (GMYC, PTP, BIN, ABGD), OTU
1299 numbers and phylogenetic status (ingroup, outgroup)

1300

1301 Table S3. Most likely substitution models for each molecular marker within each group,
1302 selected using jModelTest

1303

1304 Table S4. Likelihood scores for each diversification model for each group. Description of
1305 each model can be found in Table S1

1306 Table S5. Likelihood scores for each diversification model for each of the seven clock rate
1307 hypotheses (0.4% to 1.6% per million years) for *Channa*.

1308

1309 Table S6. Log likelihood scores and Akaike Information Criterion values for DEC, DEC+J,
1310 DIV, DIV+J, BAY and BAY+J as calculated using BioGeoBEARS for island-based and
1311 paleoriver-based geographic areas

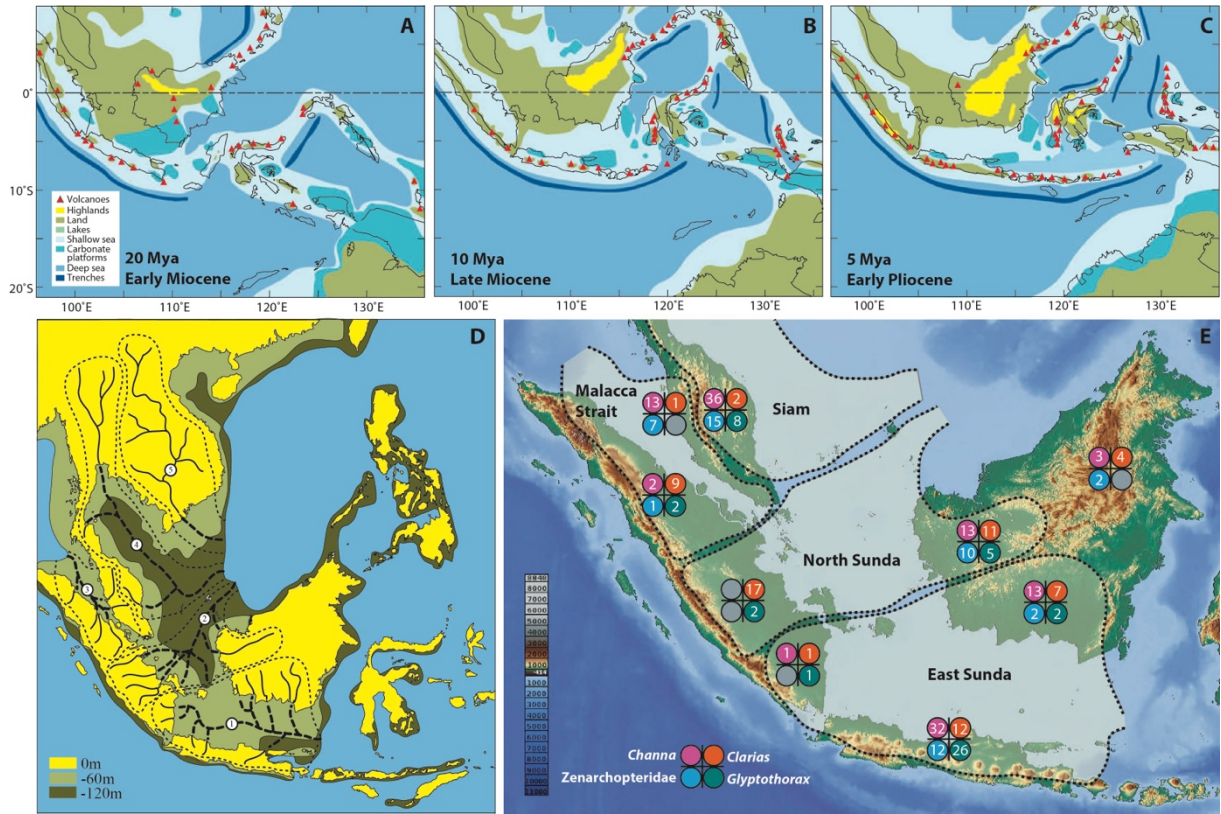


Fig. 1

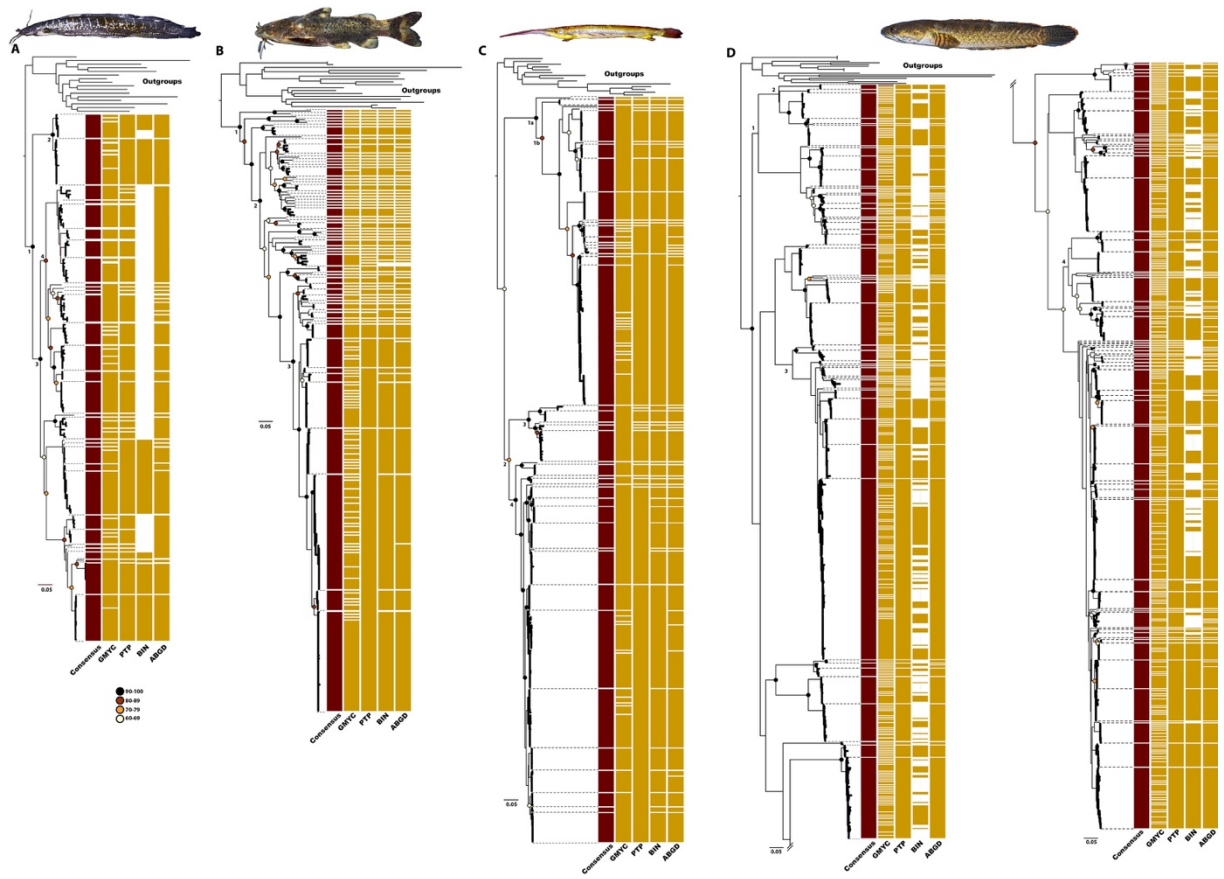


Fig. 2

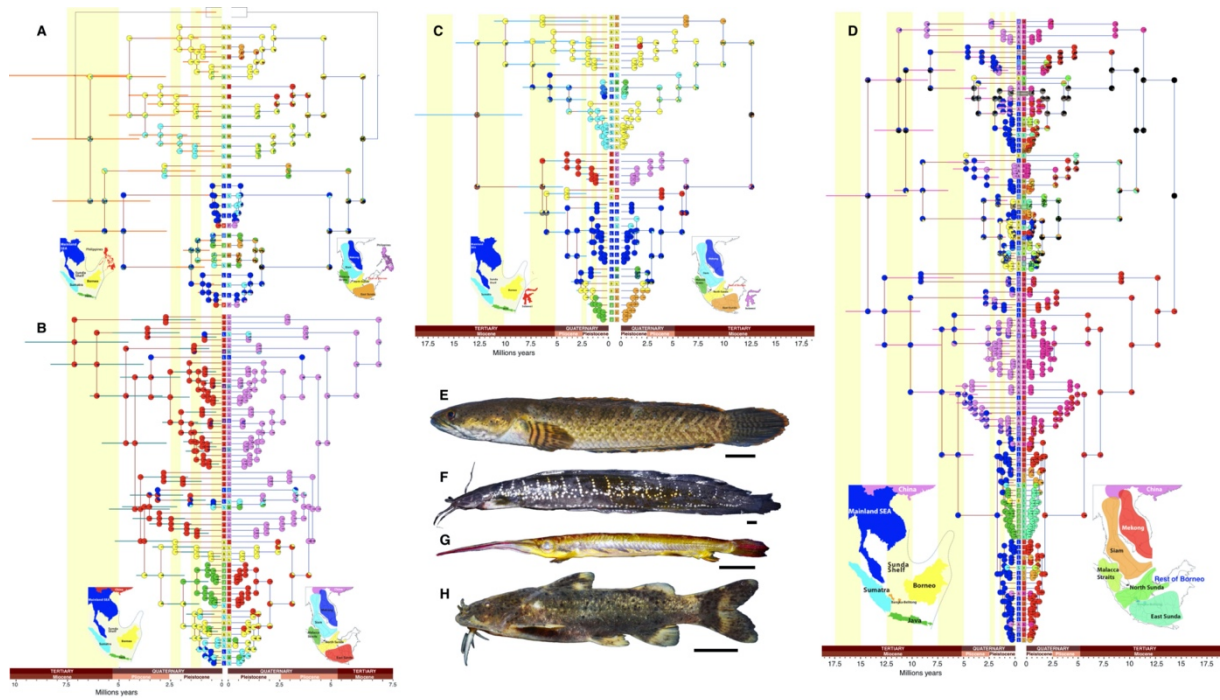


Fig. 3

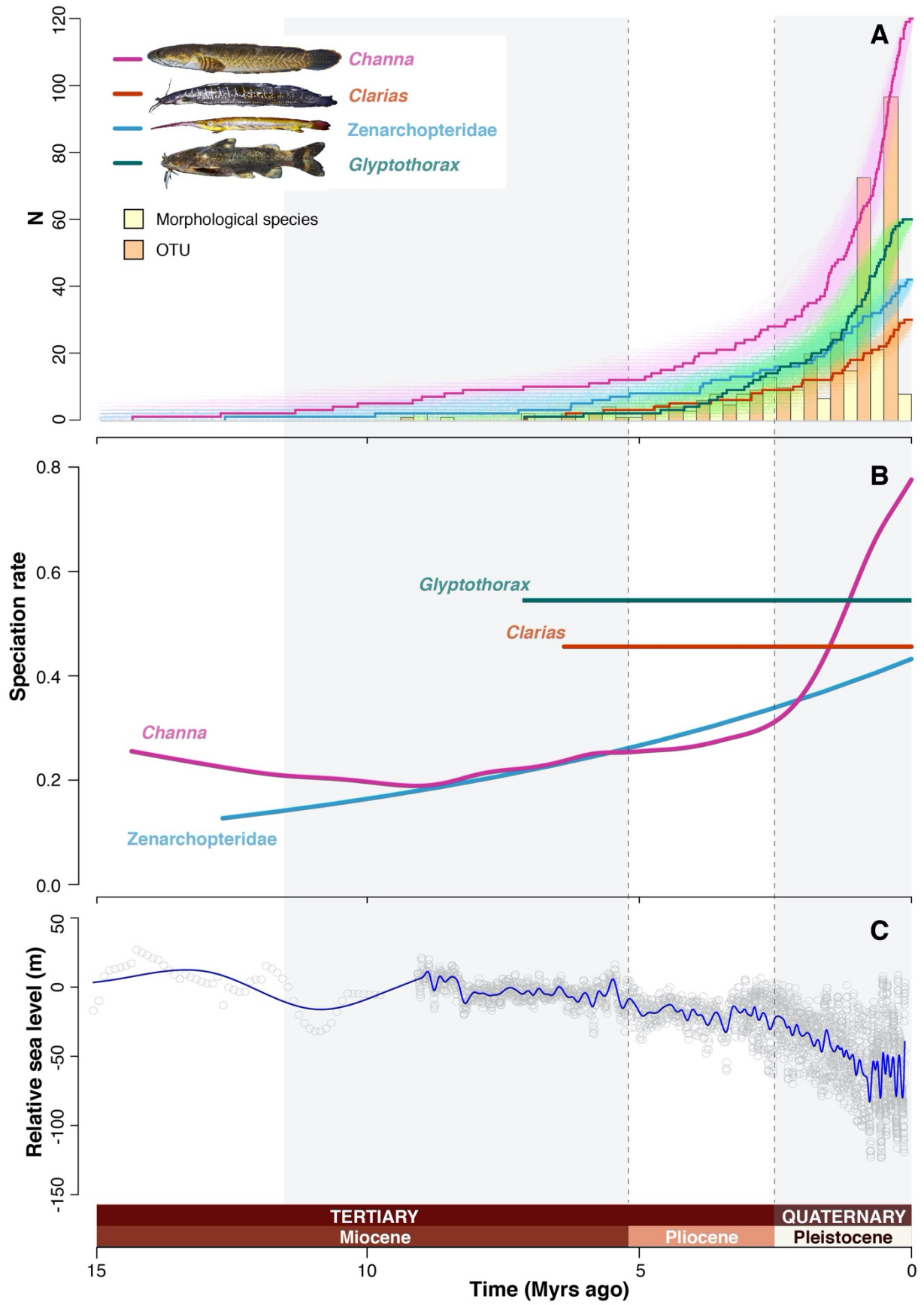


Fig. 4

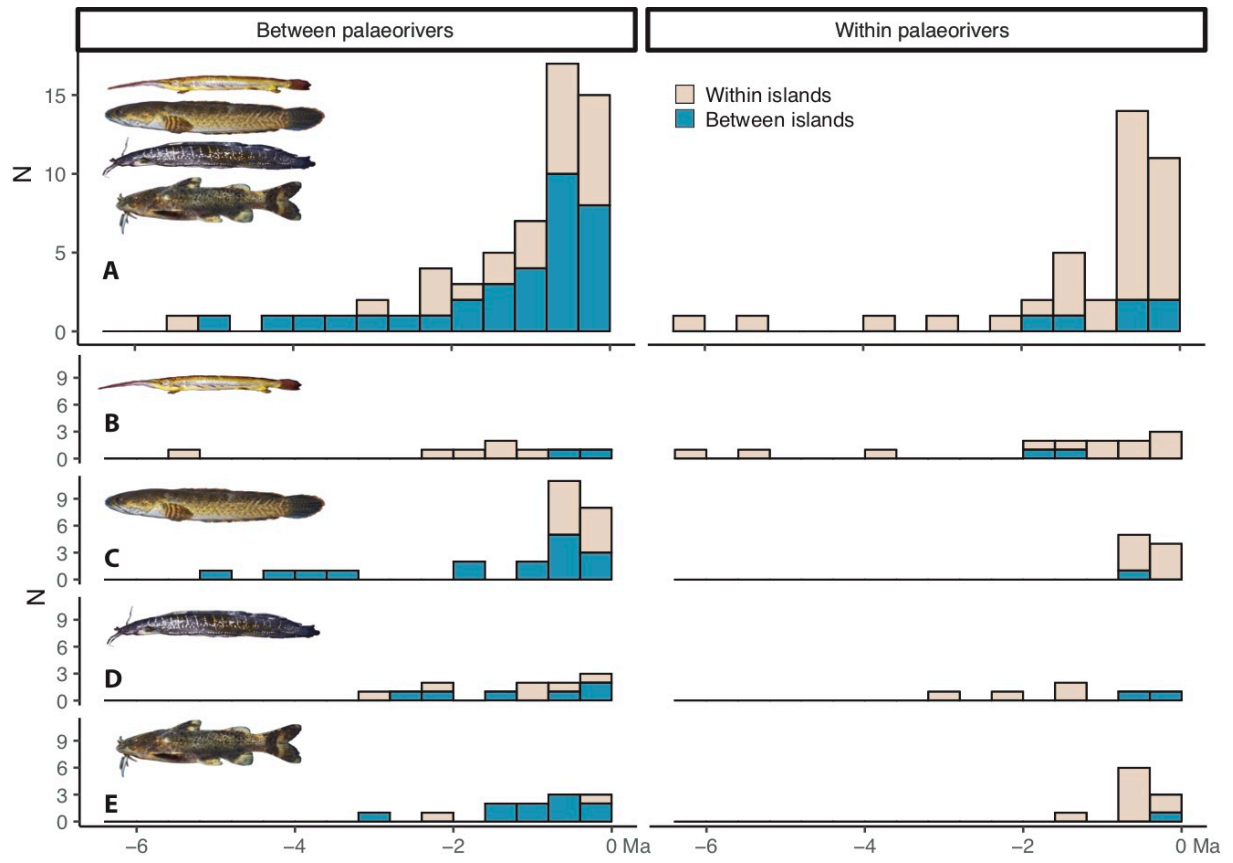


Fig. 5