



Limited dispersal and in situ diversification drive the evolutionary history of Rasborinae fishes in Sundaland

Arni Sholihah, Erwan Delrieu-trottin, Tedjo Sukmono, Hadi Dahruddin, Juliette Pouzadoux, Marie-ka Tilak, Yuli Fitriana, Jean-françois Agnèse, Fabien L. Condamine, Daisy Wowor, et al.

► To cite this version:

Arni Sholihah, Erwan Delrieu-trottin, Tedjo Sukmono, Hadi Dahruddin, Juliette Pouzadoux, et al.. Limited dispersal and in situ diversification drive the evolutionary history of Rasborinae fishes in Sundaland. *Journal of Biogeography*, 2021, 48 (9), pp.2153-2173. 10.1111/jbi.14141 . hal-03370830

HAL Id: hal-03370830

<https://hal.science/hal-03370830>

Submitted on 17 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.

Research article: Limited dispersal and *in situ* diversification drive the evolutionary history of Rasborinae fishes in Sundaland

Running title: Diversification of Sundaland's Rasborinae

Arni Sholihah^{1,2}, Erwan Delrieu-Trottin^{1,3}, Tedjo Sukmono⁴, Hadi Darhuddin⁵, Juliette Pouzadoux^{1,6}, Marie-Ka Tilak¹, Yuli Fitriana⁵, Jean-François Agnès¹, Fabien L. Condamine¹, Daisy Wowor⁵, Lukas Rüber^{7,8}, Nicolas Hubert¹

¹UMR 5554 ISEM (IRD, UM, CNRS, EPHE), Université de Montpellier, Place Eugène Bataillon, 34095 Montpellier cedex 05, France

²Institut Teknologi Bandung, School of Life Sciences and Technology, Jalan Ganesha 10, Bandung 40132, Indonesia

³Museum für Naturkunde, Leibniz-Institut für Evolutions und Biodiversitätsforschung an der Humboldt-Universität zu Berlin, Invalidenstrasse 43, Berlin 10115, Germany

⁴Universitas Jambi, Department of Biology, Jalan Lintas Jambi - Muara Bulian Km15, 36122 Jambi, Sumatra, Indonesia

⁵Division of Zoology, Research Center for Biology, Indonesian Institute of Sciences (LIPI), Jalan Raya Jakarta Bogor Km 46, Cibinong 16911, Indonesia

⁶UMR 5244 IHPE (CNRS, IFREMER, UM, UPVD), Université de Montpellier, Place Eugène Bataillon, 34095 Montpellier cedex 05, France

⁷Naturhistorisches Museum Bern, Bernastrasse 15, Bern 3005, Switzerland

⁸Aquatic Ecology and Evolution, Institute of Ecology and Evolution, University of Bern, 3012 Bern, Switzerland

Correspondence

Nicolas Hubert, UMR5554-ISEM, Université de Montpellier, Place Eugène Bataillon, 34095 Montpellier cedex 05, France.

Email : nicolas.hubert@ird.fr

Funding information:

This study was supported by IRD through annual allocations (2012-2020) and incentive funds (2012-2013), the Ministry of Foreign Affairs and International Development through a BIO-Asia grant (BIOSHOT 2016-2017) managed by Campus France, the “Institut Français

35 d'Indonésie” through a “Science et impacts” grant (2016-5758) and the European
36 Commission through a grant from the Southeast Asia – Europe Joint Funding scheme for
37 Research and Innovation program (FRESHBIO, 307943/00). AS benefited from a PhD
38 scholarship from the Indonesian Endowment Fund for Education (LPDP). The Indonesian
39 ministry of research and technology approved this study.

40

Abstract

Aim: Sea level changes have long been put forward to explain the colonization of Southeast Asian islands by freshwater aquatic organisms. We examined the relative impact of Sundaland geology since the Oligocene and of Pleistocene Eustatic Fluctuations on the mitochondrial lineage diversification of a species-rich subfamily of freshwater fishes widely distributed in Southeast Asia. We specifically tested if the expansion of exposed lands and increased island connectivity during Pleistocene low sea levels (the Paleoriver hypothesis) induced bursts of diversification.

Location: Sundaland

Taxon: Rasborinae (Actinopterygii, Cypriniformes, Danionidae)

Methods: We aggregated 1,017 cytochrome oxidase I sequences and 79 mitogenomes to delineate Molecular Operational Taxonomic Units (MOTUs) and further reconstruct a time-calibrated phylogeny of Rasborinae. Ancestral area estimations were conducted using both island and paleoriver partitioning to examine the impact of island connectivity during Pleistocene sea level changes on dispersal. Temporal trends of diversification are explored through statistical selection of best-fit models.

Results: The origin of Sundaland mitochondrial lineages is dated at ca. 33 Ma and four major clades are identified, which diversified between ca. 31 and 22 Ma. The Island of Borneo and North Sunda paleoriver are identified as the source of Sundaland Rasborinae. Geographical patterns of lineage divergence indicate that most divergence events occurred within islands and diversification under constant birth rate models are the most likely for all clades.

Conclusions: The geographical and historical context of diversification of mitochondrial lineages in Rasborinae provides little support for the Paleoriver Hypothesis. The onset of isolation of Borneo from mainland Asia triggered the initial diversification of the group (ca. 31-22 Ma). The late colonization of Java and Sumatra occurred through several independent dispersal events, poorly explained by Pleistocene sea level changes, and frequently followed by *in situ* diversification.

KEYWORDS

Ancestral areas estimation, Cryptic diversity, Dispersal pathways, Diversification models, Historical biogeography, Mitogenomes, Pleistocene climatic fluctuations, Phylogenetic reconstructions, Sea levels.

1. INTRODUCTION

Sundaland has long attracted the attention of evolutionary biologists. From his observations in the 19th century, Alfred Wallace already deduced the biological uniqueness of the “Indo-Malay Islands” compared to neighbouring continental Asia and Celebes (Wallace, 1869). It is now acknowledged that Sundaland’s diversity and endemism occur in an area where the geological history is intricate (Hall, 2009, 2013; Lohman et al., 2011). Including the Malay Peninsula and the islands of Borneo, Sumatra, Java and Bali, Sundaland emerged since the Paleocene, ca. 60 million years ago (Ma), until the middle Oligocene ca. 30 Ma, as a promontory at the southern end of Eurasia (Fig. 1A). Complex tectonic movements during the Miocene triggered the formation of Borneo between 20 and 10 Ma (Figs. 1B & 1C) and the subsequent emergence of Sumatra and Java between 10 and 5 Ma (Figs. 1C & 1D). Insular Sundaland remained partially connected until entering the Pliocene (5.33-2.58 Ma). Upon entering early Quaternary (2.6 Ma), Pleistocene Climatic Fluctuations (PCFs) led to global variations of average temperature (Westerhold et al., 2020; Zachos, Dickens, & Zeebe, 2008) and sea levels (Miller et al., 2005). Geology and Pleistocene eustasy interacted in Sundaland (Husson, Boucher, Sarr, Sepulchre, & Cahyarini, 2020; Sarr, Sepulchre, & Husson, 2019; Sathiamurthy & Voris, 2006; Voris, 2000; Woodruff, 2010). During Pleistocene glacial periods, sea level dropped between -60m and -120m and created connections between Sundaland’s islands (Fig. 1D). These exposed land areas likely had freshwater drainage systems that extended between modern islands, and this is central to the Palaeoriver Hypothesis with respect to fish species dispersal (Kottelat, Whitten, Kartikasari, & Wirjoatmodjo, 1993; Voris, 2000). Four major paleorivers occurred in Southeast Asia: (1) East Sunda, (2) North Sunda, (3) Malacca straits, and (4) Siam (Fig. 1D). These paleorivers likely impacted the dispersal of Sundaland’s freshwater biotas and their evolutionary history (de Bruyn et al., 2013).

A high proportion of Sundaland contemporary freshwater diversity corresponds to cryptic lineages with Pleistocene origins (Dahrudin et al., 2017; de Bruyn et al., 2014; Hubert, Kadarusman, et al., 2015; Hubert et al., 2019; Hutama et al., 2017; Kusuma, Ratmuangkhwang, & Kumazawa, 2016; Farhana et al., 2018; Sholihah et al., 2021, 2020). Several studies detected congruence between the distribution of some freshwater lineages and the boundaries of paleoriver watersheds (Beck et al., 2017; de Bruyn et al., 2013; Dodson, Colombani, & Ng, 1995; Tan, Jamsari, & Siti Azizah, 2012). The timing of speciation across clades during the Pleistocene indicates that PCFs may have triggered bursts of species proliferation in Sundaland through sea level changes (Fig. 2), a mechanism already suggested

in other areas (Barraclough & Nee, 2001; Cannon, Morley, & Bush, 2009; Condamine et al., 2015; Hubert & Renno, 2006; Mittelbach et al., 2007; Nores, 1999; Wiens & Donoghue, 2004). Nevertheless, other factors likely contributed to the build-up of Sundaland diversity, including: (1) pre-Pleistocene geology (Beck et al., 2017; Condamine, Toussaint, et al., 2013; de Bruyn et al., 2014; Dodson et al., 1995); (2) dynamic interactions between insular and paleoriver watershed boundaries (Brown et al., 2013; Esselstyn & Brown, 2009; Papadopoulou & Knowles, 2015a, 2015b; Sholihah et al., 2021); (3) varying dispersal abilities as a consequence of life history traits influencing abilities to colonize novel environments, and pass biotic and abiotic filters (Beck et al., 2017; Esselstyn, Timm, & Brown, 2009; Hubert, Calcagno, Etienne, & Mouquet, 2015; Logue, Mouquet, Hannes, Hillebrand, et al., 2011; Patel, Weckstein, Patane, Bates, & Aleixo, 2011; Pouyaud, Sudarto, & Paradis, 2009); and (4) habitat rearrangements during PCFs (Bird, Taylor, & Hunt, 2005; Heaney, 1992; Wurster, Rifai, Zhou, Haig, & Bird, 2019). By contrast to the Last Glacial Maximum (LGM) ca. 17,000 years ago, which housed the maximal extension of paleoriver watersheds, Sundaland is currently considered in a refugial state. This statement underlines the importance to understand the impact of PCFs on diversity patterns in Sundaland for effective conservation efforts. Sundaland is one of the most threatened biodiversity hotspots in Southeast Asia, (Hoffmann et al., 2010; Mittermeier, Turner, Larsen, Brooks, & Gascon, 2011; Myers, Mittermeier, Mittermeier, da Fonseca, & Kent, 2000), mostly due to its alarming rate of deforestation during the last decades (Gaveau et al., 2014; Imai, Furukawa, Tsujino, Kitamura, & Yumoto, 2018; Laumonier et al., 2010), in conjunction with water contamination (Garg, Hamilton, Hochard, Kresch, & Talbot, 2018) and watershed fragmentation through the development of dams for irrigation and hydroelectric powerplants (Mulligan, van Soesbergen, & Sáenz, 2020).

Freshwater fishes are tightly dependent of watershed dynamics and constitute model systems to trace historical watershed dynamics (Bernatchez & Wilson, 1998; de Bruyn et al., 2013; Durand, Persat, & Bouvet, 1999; Hubert et al., 2007). Sundaland hosts several species-rich groups that are particularly suited to explore the consequences of PCFs on freshwater diversity patterns (Hubert, Kadarusman, et al., 2015). One of these groups is the Rasborinae (Cypriniformes, Danionidae), a subfamily of iconic and highly diversified, small-size species widely distributed in Sundaland (Brittan, 1972; Dahrudin et al., 2017; Hubert et al., 2019; Kusuma et al., 2016; Liao, Kullander, & Fang, 2011; Sholihah et al., 2020; Tan & Armbruster, 2018). Rasborinae comprises eleven genera variously distributed in Asia, of which seven are endemic of Sundaland or much more diverse here than in adjacent areas

(Sholihah et al., 2020). These include *Boraras*, *Brevibora*, *Kottelatia*, *Pectenocypris*, *Rasbora*, *Trigonopoma* and *Trigonostigma* (Sholihah et al., 2020; Tan & Armbruster, 2018). The systematics of the subfamily is still confused due to the lack of robust phylogenetic hypothesis of intrarelationships (Brittan, 1972; Hui & Chian, 2020; Kottelat & Vidthayanon, 1993; Kottelat & Witte, 1999; Liao, Kullander, & Fang, 2010; Liao et al., 2011). However, the monophyly of Rasborinae is supported by both molecular and morphological characters (Saitoh et al., 2006; Stout, Tan, Lemmon, Lemmon, & Armbruster, 2016; M. Tan & Armbruster, 2018; K. L. Tang et al., 2010). Rasborinae encompasses ca. 80 species in Sundaland (Sholihah et al., 2020), representing ca. 75% of Rasborinae diversity (Eschmeyer, Fricke, & van der Laan, 2018; Froese & Pauly, 2020). Recent genetic reappraisals of Rasborinae species diversity in Sundaland confirmed species boundaries and distribution ranges, along with the recognition of a substantial amount of cryptic diversity (Dahrudin et al., 2017; Hubert et al., 2019; Sholihah et al., 2020). Most cryptic lineages identified have very narrow, non-overlapping distribution ranges, suggesting landscape fragmentation mostly contributed to generate this diversity (Sholihah et al., 2020). However, the origin and underlying evolutionary dynamics of this diversity remain largely unexplored. In particular, the historical connectivity between Sundaland and the continent during the isolation of Borneo (ca. 20-30 Ma), and the subsequent emergence of Sumatra (ca. 20-10 Ma) and Java (ca. 10-5 Ma), questions the dynamics of colonization of Sundaland by Rasborinae. Several potential scenarios can be predicted, ranging from an origin through *in situ* lineage proliferation initiated in Borneo after its isolation (ca. 20-30 Ma) to multiple immigrations events from the continent during the emergence of Sundaland islands (ca. 30-5 Ma) (Sholihah et al., 2021; Šlechtová, Musilova, Tan, Kottelat, & Bohlen, 2021).

Here, we explore the phylogenetic relationships of Sundaland's Rasborinae through mitochondrial genome skimming (Dodsworth, 2015; Straub et al., 2012), with the aim to examine the potential impacts of Sundaland geology and Pleistocene sea level changes on mitochondrial lineage proliferation of this species-rich group. We addressed the following questions: (1) Did Rasborinae diversify (a) through *in situ* radiation in Sundaland (i.e. monophyly of Sundaland Rasborinae) once Borneo isolated from Mainland (ca. 30-20 Ma) or (b) through multiple colonization events from Mainland followed by *in situ* diversification (i.e. polyphyly of Sundaland lineages) during the emergence of Sundaland (ca. 30-5 Ma)? (2) Did paleorivers serve as dispersal pathways between islands during Pleistocene low sea levels and enabling the colonization of Sundaland islands? (3) Did paleoriver watersheds prompt allopatric divergence across their boundaries? (4) Did PCF affect rates of mitochondrial

lineage diversification? Through a dense taxonomic, spatial and mitochondrial genomic sampling, Rasborinae phylogenetic intrarelationships were inferred and mitochondrial divergence events dated to explore dynamics of lineage proliferation through ancestral area estimations and statistical selection of best-fit diversification models.

2. MATERIALS AND METHODS

2.1 Analytical procedure and sampling

The species diversity of the subfamily Rasborinae has been recently revisited in Sundaland through standardized mitochondrial DNA-based approaches (Dahrudin et al., 2017; Hubert et al., 2019; Sholihah et al., 2020). Here, the objective is to take advantage of this updated cytochrome oxidase I (i.e. DNA barcode) reference library to explore the Rasborinae intrarelationships and guide taxon sampling for further mitogenome skimming (Dodsworth, 2015; Straub et al., 2012). As the earliest branching events and intrarelationships are still unknown in Rasborinae (Brittan, 1972; Liao et al., 2010), mitogenomes were first used to reconstruct a backbone phylogeny and identify major clades. Then, mitogenomes were combined to all DNA barcodes available to reconstruct phylogenetic relationships within each clade based on a dense taxon sampling. Sampling and collection management is as described in Sholihah et al. (2020). Specimens were captured using gears such as electrofishing, seine nets, cast nets and gill nets across sites that encompass the diversity of freshwater lentic and lotic habitats in Sundaland (Fig. 1). Specimens were identified following original descriptions where available, as well as monographs (Kottelat, 2013; Kottelat et al., 1993) and further validated through DNA barcodes by including records from type localities (Hubert et al., 2019; Sholihah et al., 2020). Species names were further validated using several online catalogues (Eschmeyer et al., 2018; Froese & Pauly, 2020). Specimens were photographed, individually labelled, and voucher specimens were preserved in a 5% formalin solution. Prior to fixation a fin clip or a muscle biopsy was taken and fixed separately in a 96% ethanol solution for further genetic analyses. Both tissues and voucher specimens were deposited in the national collections at the Museum Zoologicum Bogoriense (MZB), Research Center for Biology (RCB), and Indonesian Institute of Sciences (LIPI).

2.2 Mitogenomes skimming

Genomic DNA was extracted using a MINIPREP SIGMA extraction kit following manufacturer's specifications. A 651-bp segment from the 5' region of the cytochrome oxidase I gene (COI) was amplified as described in Sholihah et al. (2020). PCR

amplifications were done using the primers cocktails C_FishF1t1/C_FishR1t1 including M13 tails (Ivanova, Zemlak, Hanner, & Hébert, 2007) in a final volume of 10.0µl containing 5.0µl Buffer 2X, 3.3µl ultrapure water, 1.0µl each primer (10µM), 0.2µl enzyme Phire Hot Start II DNA polymerase (5U) and 0.5µl of DNA template (~50 ng). Amplifications were conducted as followed: initial denaturation at 98°C for 5 min followed by 30 cycles denaturation at 98°C for 5s, annealing at 56°C for 20s and extension at 72°C for 30s, followed by a final extension step at 72°C for 5 min. The PCR products were purified with ExoSap-IT (USB Corporation, Cleveland, OH, USA) and sequenced in both directions. The sequences and collateral information are available in BOLD (Ratnasingham & Hebert, 2007) in the data set DS-BIFRA (Table S1, dx.doi.org/10.5883/DS-BIFRA).

Genomic libraries for mitogenome skimming were prepared following the protocol developed by Tilak et al. (2015) for multiplexed Illumina sequencing. Genomic DNA was physically fragmented through ultrasound (35 kHz) for a duration varying between 10 and 20 min using an ultrasonic cleaning unit (Elmasonic One). We followed the Illumina library preparation procedure with blunt-end repair, adapter ligation, adapter fill-in and indexing PCR steps (13 cycles) developed by Meyer and Kircher (Meyer & Kircher, 2010). Each step was followed by a purification using SPRI bead suspensions (Agencourt® AMPure® XP), adding 1.7 volume of Agencourt® AMPure® XP reagent per volume of sample and eluted in 25µl of ultra-pure water. Quantification of DNA libraries was done with a Nanodrop ND-800 spectrophotometer (Nanodrop technologies). Indexed libraries were pooled using their relative concentrations to ensure equimolarity and a single pool was single-read sequenced (150 bp long reads) on Illumina HiSeq 2500 at MGX (Montpellier, France). Mitogenomes were then assembled on Unipro UGENE (Okonechnikov et al., 2012) by reference to the closest mitogenome available among the 10 Rasborinae mitogenomes available in GenBank. Complete mitogenomes were then annotated using the online tool MitoAnnotator (Iwasaki et al., 2013) available at mitofish.aori.u-tokyo.ac.jp. Annotated mitogenomes are accessible in GenBank (Table S1).

2.3 Reconstructing a backbone phylogeny of Rasborinae with mitogenomes

The Most Recent Common Ancestor (MRCA) of Rasborinae has been previously estimated at 43 Ma (Betancur-R et al., 2017). Thus, tRNAs and control region were trimmed for phylogenetic reconstructions due to their fast substitution rates, and likely high levels of homoplasy. Protein and ribosomal RNA (rRNA) coding regions were retained and

individually partitioned in subsequent phylogenetic reconstructions. First, a maximum likelihood (ML) tree was reconstructed using a partitioned model for each protein and rRNA coding regions with a GTR+I+ Γ model as implemented in RAxML-HPC Blackbox (Miller, Pfeiffer, & Schwartz, 2010) with RAxML 8 (Stamatakis, 2014). Topological support was estimated with 5,000 non-parametric bootstrap replicates. Second, a calibrated tree was reconstructed using Bayesian inferences as implemented in BEAST 2.6.2 (Bouckaert et al., 2014; Heled & Drummond, 2010). The most likely substitution models were jointly determined for all partitions using ModelFinder (Kalyaanamoorthy, Minh, Wong, von Haeseler, & Jermini, 2017) as implemented in the IQTREE online webserver (Nguyen, Schmidt, Von Haeseler, & Minh, 2015) at <http://iqtree.cibiv.univie.ac.at>. The selected models were further used to conduct a Bayesian partitioned analysis based on a Yule model (uniform birth rate), relaxed clock with log normal distribution, and standardized site models as implemented in the SSM package in BEAST 2.6.2. Two Monte Carlo Markov Chains (MCMC) of 50 million generations (burnin of 10%) were conducted to check for convergence and to check that ESS estimates reached 200 using Tracer 1.7.1 (Drummond, Suchard, Xie, & Rambault, 2012). Two clock rates were jointly estimated for rRNA and protein-coding regions along tree topology. MCMCs were initiated with a 0.3% of divergence per million years (Myrs) for rRNA (Hardman & Lundberg, 2006; Orti, 1997) and 1.2% per Myrs for protein-coding regions (Bermingham, McCafferty, & Martin, 1997). The two runs were then combined with LogCombiner 2.6.2 (Bouckaert et al., 2014) and the maximum clade credibility tree, median age estimates and corresponding 95% highest posterior density (HPD) were summarized using TreeAnnotator 2.6.2 (Bouckaert et al., 2014). This analysis was repeated with a birth-death model of diversification to assess the potential influence of the branching process prior on divergence time estimates (Ritchie, Lo, & Ho, 2017). Both ML and Bayesian inferences were rooted using an assortment of mitogenomes (Table S1) available for several closely related subfamilies of Danionidae as well as other Cypriniformes families following previously published phylogenetic hypotheses (Betancur-R et al., 2017; Chang, Rabosky, Smith, & Alfaro, 2019; Rabosky et al., 2018; K. L. Tang et al., 2010).

2.4 DNA barcodes, mitochondrial lineage delimitation and clade-specific phylogenetic inferences

Once major clades were identified within the subfamily, all DNA barcodes available from previous studies were compiled (Table S1). DNA barcode sequences were selected according to a preliminary screening of their phylogenetic affinities and robustness of the

phylogenetic inferences. A Neighbour Joining (NJ) was first reconstructed for the 1,097 DNA barcodes from Sholihah et al. (2020) and branching support was estimated through 5,000 bootstrap replicates using PAUP 4.0a (Swofford, 2001). Only DNA barcode records related to species with mitogenomes available, with bootstrap proportions (BP) above 80% were retained. Genetic delimitation of species follows the protocol described in Sholihah et al. (2020). Four different sequence-based methods of species delimitation were used to delimitate Molecular Operational Taxonomic Units (MOTUs) (Blaxter et al., 2005) using the 1,097 DNA barcodes dataset. These methods were: (1) Refined Single Linkage (RESL) as implemented in BOLD and used to generate Barcode Index Numbers (BIN) (Sujeevan Ratnasingham & Hebert, 2013), (2) Automatic Barcode Gap Discovery (ABGD) (Puillandre, Lambert, Brouillet, & Achaz, 2012), (3) Poisson Tree Process (PTP) in its multiple rates version (mPTP) as implemented in the stand-alone software mptp_0.2.3 (Zhang, Kapli, Pavlidis, & Stamatakis, 2013), and (4) General Mixed Yule-Coalescent (GMYC) in its multiple rate version (mGMYC) as implemented in the R package 'Splits' 1.0-19 (Fujiwasa & Barraclough, 2013). RESL and ABGD used DNA alignments as input files, while a ML tree was used for mPTP and a Bayesian Chronogram based on a strict-clock model using a 1.2% of genetic distance per Myrs for mGMYC. The ML tree for mPTP was reconstructed using RAxML 8 using a GTR+ Γ substitution model (Stamatakis, 2014) and the ultrametric and fully resolved tree for mGMYC was reconstructed using BEAST 2.6.2 with two independent Markov chains of 50 million generations each including a Yule pure birth model tree prior, a strict-clock model and a GTR+I+ Γ substitution model. Both runs were combined using LogCombiner 2.6.2 and the maximum clade credibility tree was constructed using TreeAnnotator 2.6.2. Duplicated haplotypes were pruned for further species delimitation analyses.

Once DNA barcodes were selected and MOTUs delimited, DNA barcode and mitogenome alignments were concatenated for each of the major clades identified. Phylogenetic relationships within clades were reconstructed using the Bayesian analysis implemented in the StarBEAST2 package (Ogilvie, Bouckaert, & Drummond, 2017) from the BEAST 2.6.2 suite (Bouckaert et al., 2014). This approach implements a mixed-model including a coalescent component within MOTU and a diversification component between MOTU that allows accounting for variations of substitution rates within and between MOTUs (Ho & Larson, 2006; Ritchie et al., 2017). StarBEAST2 jointly reconstruct and optimize gene trees and MOTU trees, and as such requires the designation of MOTUs, which were determined using the majority rule consensus of the lineage delimitation analyses. After

preliminary analyses, phylogenetic reconstructions were conducted using a single partition including rRNA and protein-coding regions, GTR+I+ Γ substitution model, uncorrelated log-normal species tree model (UCLN), and MCMCs of 60 million generations. Age of each clade MRCA estimated from the backbone phylogeny was used as a calibration point with a normal distribution and a sigma of 1.0. Clock rate was estimated and an initial value of 0.8 % of divergence per Myrs was used according to the initial BEAST 2.6.2 analysis of the mitogenome dataset. Independent runs were combined using LogCombiner 2.6.2 (Bouckaert et al., 2014). Gene and MOTU maximum clade credibility trees, median age estimates and corresponding 95% HPD were summarized using TreeAnnotator 2.6.2 (Bouckaert et al., 2014). All phylogenetic reconstructions were performed using CIPRES (Miller et al., 2010), and phylogenetic trees and alignments have been deposited in Dryad (doi.org/10.5061/dryad.tb2rbp00g).

2.5 Diversification rates estimation

Lineages through time (LTT) were plotted using the MOTU trees of each major clades with the R-package ‘ape’ (Paradis & Schliep, 2019). Confidence intervals were computed using 1,000 dated trees sampled along the StarBEAST2 MCMC. To test the impact of past environmental dynamics on diversification, we relied on a ML framework with five diversification models (constant-rate, time-dependent, temperature-dependent, sea-level-dependent, and diversity-dependent models) and their variants (Condamine, Rolland, & Morlon, 2019, 2013). In total, we fitted 17 diversification models (Table S2) using the R-packages ‘RPANDA’ 1.3 (Morlon et al., 2016) and ‘DDD’ 3.7 (Etienne et al., 2012). We accounted for potential missing lineages in the phylogeny in the form of global sampling fraction, i.e. the ratio of sampled lineages diversity over the total described lineages, and ran these analyses for sampling fractions of 100% and 90%. In ‘RPANDA’, we first fitted two constant-rate models as initial references, namely: BCST (speciation rate constant through time with no extinction) and BCSTD CST (speciation and extinction rates constant through time). Second, we fitted four time-dependent models: BtimeVar (speciation rate varying through time with no extinction), BtimeVarDCST (speciation rate varying through time with constant extinction), BCSTDtimeVar (constant speciation and extinction rate varying through time) and BtimeVarDtimeVar (both speciation and extinction rates varying through time). Lastly, we fitted eight models with speciation and extinction rates varying according to external environmental variables: BtemperatureVar and Bsea-levelVar (speciation varying in

function of the environmental variable), BtemperatureVarDCST and Bsea-levelVarDCST (speciation varying in function of the variable with constant extinction rate), BCSTDtemperatureVar and BCSTDsea-levelVar (extinction rate varying in function of the variable and constant speciation rate) and BtemperatureVarDtemperatureVar and Bsea-levelVarDsea-levelVar (both speciation and extinction rates varying in function of the variable).

Exponential dependence with time, temperature or sea level and diversification rates are chosen for its robustness and flexibility depending on the strength and direction of the dependence to the fitted variable. Speciation (λ) and extinction (μ) rates are parameterized as follows. When λ and μ are exponential functions of sea level (S) through time (t), the equations are $\lambda(S_{(t)}) = \lambda_0 \times e^{\alpha S(t)}$ and $\mu(S_{(t)}) = \mu_0 \times e^{\beta S(t)}$, where λ_0 and μ_0 are respectively the expected λ and μ at $S = 0$ meter, while α and β are coefficients that measure the strength and the sign of the relationship with sea level (e.g. $\alpha > 0$ and $\beta > 0$ respectively indicate λ and μ increase with sea level high stands). Similar parameterisation can be used for exponential relationship between temperature (T) through time (t) and λ as well as μ rates in which $\lambda(T_{(t)}) = \lambda_0 + \alpha T_{(t)}$ or $\lambda(T_{(t)}) = \lambda_0 \times e^{\alpha T(t)}$, where $T(t)$ is the temperature at time t and λ_0 is speciation rate at $T = 0^\circ\text{C}$.

The last three diversity-dependent ML models were fitted using ‘DDD’ package in which λ and μ vary as linear functions of number of lineages within each clade (Etienne et al., 2012). The diversity-dependent models are parameterised by λ_0 , μ_0 (respectively indicating λ and μ at the absence of competing lineage), K (carrying capacity, representing asymptotic clade size). All λ and μ were constrained to be positive. The results were then compiled with the previous 14 models in ‘RPANDA’ and all are compared using corrected Akaike Information Criterion (AICc) and Akaike weights (AIC ω). The model with the lowest AICc and highest AIC ω was considered as the best fitting model for the phylogeny.

2.6 Ancestral areas estimation

To explore dispersal and vicariance history based on the Paleoriver Hypothesis, we reconstructed ancestral distribution of Sundaland Rasborinae using the R-package ‘BioGeoBEARS’ 1.1.2 (Matzke, 2014; Matzke, 2013) based on the StarBEAST2 MOTU trees. Species presence/absence were compiled (Table S1) for two sets of geographical delimitation based on: (1) paleoriver and (2) contemporary island boundaries. Then, geographical patterns of divergence were recorded as follows: (1) no dispersal, sister lineages

co-occur within the same paleoriver and the same island; (2) dispersal between islands within a paleoriver, sister lineages are alternatively distributed on different islands within the same paleoriver; (3) dispersal between paleorivers within the same island, sister lineages are alternatively distributed on different paleorivers within the same islands; and (4) dispersal between islands and between paleorivers, sister lineages are alternatively distributed on different paleorivers and different islands. Ancestral area estimations involving paleorivers were based on the following geographical areas (Fig. 1): (1) the Malacca Straits paleoriver, (2) the East Sunda paleoriver, (3) the North Sunda paleoriver, (4) the Siam paleoriver, (5) the Northern Borneo river system, and (6) the Mekong river system. On the other hand, estimations based on insular delimitation followed geographical divisions of: (1) Sumatra-Bangka-Belitung, (2) Java-Bali-Lombok, (3) Borneo, and (4) Mainland Southeast Asia. For analytical requirements, only MOTUs with known localities that could be associated to both islands and palaeodrainages were used.

Inferences of the ancestral areas using ‘BioGeoBEARS’ were conducted using six alternative models including dispersal-extinction cladogenesis (DEC), DEC+J, dispersal-vicariance analysis ML-version (DIVALIKE), DIVALIKE+J, Bayesian biogeographical inference model (BAYAREALIKE) and BAYAREALIKE+J (Matzke, 2014; Van Dam & Matzke, 2016). The inclusion of the parameter J has been recently criticized from a conceptual and statistical perspective (Ree & Sanmartín, 2018). The concept of jumping dispersal has been developed for insular systems to account for the settlement of a new lineage established by colonization without an intermediate widespread ancestor (Clark et al., 2008; Ree & Sanmartín, 2018). Considering the biogeographical scenario of Sundaland and the insularity of the system, jumping dispersal cannot be discarded *a priori* from a conceptual perspective and several studies have previously highlighted the importance of jumping dispersal in insular systems (Beck et al., 2017; Condamine et al., 2015; Cowie & Holland, 2006; de Bruyn et al., 2013; Hendriks, Alciatore, Schilthuizen, & Etienne, 2019; Sholihah et al., 2021). Models of ancestral area estimation including the J parameters were thus considered here and the best-fit model was estimated using the AICc.

3. RESULTS

3.1 Mitogenomes skimming

Illumina sequencing yielded a number of reads ranging between 4.5 and 12 million per genomic libraries. A total of 58 new mitogenomes were successfully assembled with a minimum coverage of 10X for four genera including three *Brevibora*, one *Pectenocypris*, 45

Rasbora and seven *Trigonopoma*. Mitogenomes were ca. 16,500 bp long on average and include the coding regions of two rRNA (12S, 16S), 22 tRNA, 13 protein-coding genes and the Control Region (CR). In addition, 10 Rasborinae mitogenomes were retrieved from GenBank including the genera *Amblypharyngodon*, *Horadandia*, *Rasboroides*, *Rasbora* and *Boraras* (Table S1). These 10 mitogenomes were used as reference genomes for assembly and further phylogenetic reconstructions. An additional set of 11 mitogenomes was retrieved from GenBank of Danionidae subfamilies and Cypriniformes families, which were used as outgroups (Table S1). For phylogenetic reconstruction, tRNA coding regions and CR were trimmed. The final alignment included 79 mitogenomes and 13,898 bp consisting of 2,518 bp of rRNA and the 13 protein-coding regions including ND1 (975 bp), ND2 (1,045 bp), COI (1551 bp), COII (691 bp), ATP8 (1655 bp), ATP6 (673 bp), COIII (785 bp), ND3 (349 bp), ND4L (297 bp), ND4 (1,375 bp), ND5 (1,822 bp), ND6 (518 bp), and Cytb (1,134 bp).

3.2 Reconstructing the backbone phylogeny of Rasborinae

Phylogenetic reconstructions based on the 79 mitogenomes are well supported with most internal branching events displaying posterior probabilities (PP) of 100% for the Bayesian inference (Fig. 3A) and Bootstrap Proportions (BP) above 90% for the ML reconstruction (Fig. 3B). Bayesian and ML topologies are congruent with continental Asian Rasborinae (*Amblypharyngodon*, *Horadandia* and *Rasboroides*) corresponding to the earliest branching events in Rasborinae tree and Sundaland lineages constituting a monophyletic group (Fig. 3). Four major clades are identified within Sundaland: (1) Clade I including the genera *Pectenocypris*, *Boraras*, *Trigonopoma* and *Rasbora kalbarensis*; (2) Clade II including some *Rasbora* species; (3) Clade III including some *Rasbora* species; and (4) Clade IV including some *Rasbora* species and the genera *Brevibora* and *Trigonostigma*. The MRCA of Rasborinae dates back to ca. 46 Ma (95% HPD = 40-51 Ma, Eocene), while the MRCA of the four Sundaland clades is dated at ca. 33 Ma (95% HPD = 30-37 Ma, Oligocene-Eocene transition). The four clades have varying age estimates with the MRCA of Clade I dated at ca. 31 Ma, (95% HPD = 27-34 Ma, Oligocene), the MRCA of Clade II at ca. 28 Ma (95% HPD = 24-31 Ma, Oligocene), the MRCA of Clade III at ca. 26 Ma (95% HPD = 23-30 Ma, Oligocene), and the MRCA of Clade IV at ca. 22 Ma (95% HPD = 19.36-24.86 Ma, early Miocene). Estimated clock rates ranged between 0.31% per Myrs, with a variance of 0.0017%, for rRNA coding regions and 0.74% per Myrs, with a variance of 0.0048%, for protein-coding regions. Divergence time estimates were very similar with a birth-death model

(Fig. S1). The MRCA of Sundaland Rasborinae is dated at ca. 33 Ma (95% HPD = 30-37 Ma), and divergence time estimates of descending nodes yielded very similar values.

3.3 Mitochondrial lineage delimitation and clade-specific phylogenetic inferences

The four clades have a substantial proportion of cryptic diversity (Fig. 4) as previously reported (Sholihah et al., 2020). COI sequence selection based on 80% BP threshold yielded 1,017 sequences, which together with the 66 Rasborinae mitogenomes sum up to 1,083 sequences for 71 nominal species, 157 MOTUs and 10 genera (Table S1). ML and Bayesian gene trees are congruent for all clades (Fig. 4). Most internal branching events are well supported in the ML gene trees (Figs. 4A, 4C, 4E, 4G), except among some *Trigonopoma* MOTUs within Clade I (Fig. 4A), and the most derived MOTUs of Clade IV (Fig. 4G). Bayesian gene trees were more supported for all clades (Figs. 4B, 4D, 4F, 4H) with most internal branching events supported by PP>0.9. The estimated mitogenome clock rate was 0.55% per Myrs for Clade I, 0.64% per Myrs for Clade II, 0.51% per Myrs for Clade III, and 0.56% per Myrs for Clade IV. According to the backbone phylogeny analysis and subsequent MRCA age calibrations, several clades are identified and labelled as clades I-IV. In clade I, three subclades are observed: (1) *Pectenocypris* spp., with an MRCA dated at ca. 8 Ma (95% HPD = 4-13 Ma), (2) *Boraras* spp., with a MRCA dated at ca. 24 Ma (95% HPD = 19-28 Ma), and (3) *Trigonopoma* spp., including *Kottelatia britani* and *Rasbora kalbarensis*, with a MRCA dated at ca. 26 Ma (95% HPD = 24-27 Ma). In clade II, two subclades are observed: (1) *R. lacrimula*, *R. semilineata*, *R. hubbsi*, *R. rutteni* and *R. sarawakensis*, with a MRCA dated at ca. 16 Ma (95% HPD = 14-18 Ma), and (2) all remaining species within clade II, with a MRCA dated at ca. 25 Ma (95% HPD = 24-27 Ma). In clade III, two clades can be identified: (1) *R. cephalotaenia* MOTUs, with a MRCA dated at ca. 8 Ma (95% HPD = 5-11 Ma), and (2) all remaining species within clade III, with a MRCA dated at ca. 21 Ma (95% HPD = 19-24 Ma). In clade IV, three subclades can be identified: (1) *Rasbora tornieri* and *R. dusonesis*, with a MRCA dated at ca. 17 Ma (95% HPD = 12-21 Ma), (2) *R. myersi*, *R. borapetensis*, *R. sumatrana* and *R. argyrotaenia*, with a MRCA dated at ca. 15 Ma (95% HPD = 13-26 Ma), and (3) all other *Rasbora* species within the clade, starting from *R. trilineata*, with a MRCA dated at ca. 19 Ma (95% HPD = 18-20 Ma).

3.4 Diversification rates estimation

MOTUs age estimates highlight that nearly half have Pleistocene origins (47.7%) by contrast with only 10.3% of nominal species (Fig. S2). However, this proportion varies

among clades with 55% of MOTUs with a Pleistocene origin in Clade IV, 50% in Clade III, 41% in Clade I and 36% in Clade II. Best-fit diversification models indicate no clade-specific patterns of diversification (Table S3). The constant speciation rate without extinction (BCST) model is the most likely for all clades (AIC ω of 0.23 for Clades I and II, 0.262 for Clade III, and 0.22 for Clade IV), with both sampling fractions used (Table 1, Table S3). Speciation rates are similar for Clades I, II and III with $\lambda = 0.0898, 0.108, 0.1195$ events/Myr/lineage, respectively (Table 1, Fig. S2). Clade IV, however, shows a higher rate of speciation with $\lambda = 0.147$ (Table 1, Fig. S2).

3.5 Ancestral areas estimation

A total of 139 MOTUs from 60 nominal species with known geographical distribution were used for ancestral area estimations for the four clades (Fig. 5, Fig. S4). The most likely biogeographical models include the J parameter for all clades and both island-based and paleoriver-based geographical partitioning (Table S4). DEC+J is the most likely model in most cases, except for Clade III with the paleoriver-based partitioning supporting the DIVALIKE+J model. Ancestral area estimations point to Borneo as the most likely origin of Clades II and IV and subclades of Clade II, with a high probability of Bornean ancestry of their MRCAs (Figs. 5B, 5C, 5D). The insular ancestry of Clade I is not resolved, but the North Sundaland paleoriver is identified as the most likely origin of Clade I (Fig. 5A) as well as Clades II and IV (Figs. 5B, 5D). Ancestral areas estimation indicates a recent colonization of Java during the Pliocene (ca. 4 Ma) by Clade IV (Fig. 5D). *In situ* diversification is observed in all clades, particularly in Borneo (island-based analyses) and North Sunda (in paleoriver-based analyses) as exemplified by the upper group of Clade II (*Rasbora lacrimula*, *R. hubsii*, *R. semilineata*, *R. sarawakensis*, *R. rutteni*) from Borneo (ca. 16 Ma, 95% HPD = 14-18 Ma), as well as *Trigonopoma gracile* MOTU144-MOTU148 (ca. 6 Ma, 95% HPD = 3-9 Ma) and *Trigonopoma pauciperforatum* MOTU154-MOTU157 (ca. 4 Ma, 95% HPD = 2-6 Ma) from North Sunda (Fig. 5).

Several dispersal pathways are identified between Sundaland and Asia mainland, and within Sundaland (Table 2, Figs. 5 & 6). Between 31-17.5 Ma, two pathways are identified: (a) between Borneo and Asia mainland, and (b) Borneo and Sumatra. Between 17.5-7.5 Ma, four pathways are identified: (a) between Borneo and Asia mainland, (b) between Borneo and Sumatra, (c) between Sumatra and Asia, and (d) between North and East Sunda paleorivers within Borneo. Between 7.5-0 Ma, eight pathways are identified: (a) between Borneo and

Asia mainland, (b) between Borneo and Sumatra, (c) between Sumatra and Asia mainland, (d) between North and East Sunda paleorivers within Borneo, (e) between North Sunda and Malacca paleorivers within Sumatra, (f) between Sumatra (North Sunda paleoriver) and Asia mainland, (g) between Sumatra and Java, and (h) between Borneo and Java. Several burst of *in situ* mitochondrial lineage diversification are detected within Borneo since the Early Miocene, and within Sumatra and Java since the early Pliocene (Fig. 6).

In total, most mitochondrial lineage diversification events are observed within islands with 59% and between paleorivers with 54% (Table 3, Fig. S3). Most divergence events are associated to dispersal as 66% involve either different islands or different paleorivers, while 34% occurred within the same island and the same paleoriver (Table 2). However, these trends vary through time. Most lineage divergence between islands occurs during the last 5 Myrs and is mostly associated to dispersal between paleorivers (Fig. S3A). Along the same line, *in situ* divergence within paleoriver occurs mostly within island during the last 5 Myrs (Fig. S3A). Geographical patterns of divergence vary among clades with a spectacular dominance of divergence within islands for Clade II, most frequently occurring within the same paleoriver (Table 2, Fig. S3C). Proportions are more balanced for other clades (Table 2), with most divergence events within paleorivers occurring within islands (Fig. S3).

4. DISCUSSION

4.1 Sundaland biogeography

Our phylogenetic and biogeographic reconstructions support the hypothesis of mostly *in situ* diversity build-up of Rasborinae within Sundaland, with limited connections between Sundaland and Asia mainland, during a timeframe consistent with geological reconstructions in Southeast Asia (Hall, 2009; Hall, Cottam, & Wilson, 2011; Lohman et al., 2011). Rasborinae lineages from mainland Asia are associated with the earliest branching events at the root of the Rasborinae phylogenetic reconstructions. Our divergence time estimate of Sundaland clades indicates that the four Sundaland lineages started to diversify between ca. 31-22 Ma, during the isolation of Borneo from mainland Asia (Fig. 1). These age estimates are consistent with a single origin of Sundaland Rasborinae through *in situ* diversification as evidenced by the phylogenetic reconstructions supporting their monophyly. Ancestral area estimations further confirm this scenario as Borneo is inferred as the most likely origin of most Sundaland clades (Fig. 5) and also the only island supporting early Miocene *in situ* diversification (Fig. 6A: 1). These results suggest Rasborinae of Sundaland originate from mainland Asia in the Oligocene, when Borneo was still part of mainland Asia, and

diversification of the four clades started during the isolation of Borneo. This timeframe supports the pre-Pleistocene colonisation of Sundaland by freshwater fishes (de Bruyn et al., 2013; Dodson et al., 1995; Sholihah et al., 2021), as previously reported in insular Southeast Asia (Brown et al., 2013), followed by subsequent dispersal and divergence during the emergence of Sumatra and Java (Fig. 6).

Our biogeographic estimations further highlight the importance of the North Sunda paleoriver during the initial freshwater fish diversification in Sundaland (de Bruyn et al. 2013, 2014). The North Sunda paleoriver is the most likely centre of origin of most Rasborinae clades (Fig. 4) and also the part of Borneo involved in the highest number of dispersal pathways during the late Miocene (Fig. 6B: a, b and d) and early Pliocene (Fig. 6C: a, b and d). The diversification of Rasborinae mitochondrial lineages further followed Sundaland's geological history with various colonization scenarios for Sumatra and Java. Java is the youngest of Sundaland islands, with a separation that occurred during the last 5 Myrs (Fig. 1), which is consistent with our divergence time estimates as none of the MOTUs endemic of Java are older than ca. 4 Myrs (*Rasbora* sp. 2). However, our reconstructions indicate that colonization of Java likely results from at least two distinct immigration events (Fig. 6C: g, h) followed by *in situ* diversification (Fig. 6C: 1). Immigration followed by *in situ* lineage diversification has been previously suggested for freshwater fishes in Java, including *Rasbora* (Hubert et al., 2019; Hutama et al., 2017; Kusuma et al., 2016). These divergence events are likely resulting from the intense volcanic activity in Java, which formed several volcanic arches and fragmented rivers into multiple, small and confined watersheds during the Pliocene and Pleistocene. Java is the only island of Sundaland with a predominant influence of volcanic activity during its emergence (Lohman et al., 2011). These rugged aquatic landscapes likely fragmented ancestral lineages, creating local radiations within the island (Nguyen et al. 2008; Pouyaud et al. 2009; Hubert et al. 2015, 2019; Kusuma et al. 2016; Dahrudin et al. 2017; Hutama et al. 2017).

This scenario in Java contrasts with Sumatra, which biogeographic estimations suggest a more ancient and intricate scenario of colonization (Fig. 6). Two ancient lineages are detected in Sumatra: *Rasbora kalbarensis* diverged ca. 23 Ma, and *R. jacobsoni* diverged ca. 20 Ma. These lineages occur in the North Sunda or Malacca paleorivers (Fig. 5). The age estimates of these species contrast with most lineages in Sumatra, not exceeding ca. 13 Myrs and mostly tracing back to the Miocene-Pliocene transition ca. 5 Ma. These contrasted patterns suggest two waves of colonization during: (1) the onset of Borneo isolation and the emergence of the first volcanic arches in Sumatra during the early Miocene (Fig. 6A: b), and

(2) the final stage of isolation of Sumatra, when land bridges were still connecting the Southern tip of Sumatra, Borneo and West Java during the late Miocene (Fig. 6B: b, c). Cases of *in situ* diversification are also detected in Sumatra during the Pliocene (Fig. 6C: 2).

4.2 Dispersal and Pleistocene palaeoenvironments

Our ancestral area estimations and modelling of lineage divergence and dispersal indicate jump dispersal is common in Rasborinae, a result supported by the significant increase of likelihood scores and Akaike weights when the *J* parameter is included (Table S4). Multiple cases of trans-island dispersal have been reported among Sundaland freshwater fishes (Adamson, Hurwood, & Mather, 2010; Beck et al., 2017; de Bruyn et al., 2013; Lim et al., 2016; Farhana et al., 2018; Pouyaud et al., 2009; Sholihah et al., 2021; Tan & Lim, 2013). However, our reconstructions question Pleistocene sea level changes as a main driver of dispersal between islands. During glacial maxima, sea levels dropped (Miller et al., 2005). The shallow Java Sea has given way to exposed lands, which likely had freshwater drainage systems that extended between modern islands (Voris, 2000; Woodruff, 2010). These paleoriver systems are expected to ease dispersal between islands and promote divergence (Fig. 2). This prediction received little support here. Most mitochondrial divergence events occurred within islands (Table 3) and most divergence events between islands occurred between paleorivers (Fig. S3). The predominance of divergence within islands is likely linked to Rasborinae ecology. Most species in Clades I, II, and III are small-size species inhabiting forested streams and peat swamps (Kottelat, Whitten, Kartikasari, & Wirjoatmodjo, 1993; Sholihah et al., 2020). These species are forest-dependent taxa with limited dispersal abilities.

During glacial maxima, climate was cooler and drier. Savanna and seasonal forest corridors expanded through the interior of Sundaland (Bird et al., 2005; Heaney, 1992), enhancing microclimates and diversity of freshwater habitats along the inter-island channels (Bird et al., 2005; Gorog, Sinaga, & Engstrom, 2004; Heaney, 1992; Pouyaud et al., 2009; Wurster et al., 2019). While documented for terrestrial organisms, divergence pattern of Clade II suggests that vegetational changes during glacial maxima also limited dispersal for aquatic, forest-associated organisms. Clade II further exemplifies the intricate interactions between palaeoenvironments and dispersal. Habitat specificity toward forest habitats and peat swamps may have limited dispersal between islands despite the availability of freshwater corridors. These results highlight the dynamic interactions between palaeoenvironments and Pleistocene sea level changes (Esselstyn and Brown 2009; Brown et al. 2013; Papadopoulos and Knowles 2015a, 2015b).

4.3 Macroevolutionary drivers of diversification

Pleistocene climatic fluctuations have been frequently invoked to account for Pleistocene increased rates of diversification (Mittelbach et al., 2007; Weir & Schluter, 2007). In Sundaland, Pleistocene sea level changes are predicted to induce cycle of dispersal during glacial time and vicariance during interglacial periods (Kottelat et al., 1993; Voris, 2000; Woodruff, 2010). An amplification of sea-level fluctuations during the Pleistocene has been observed (Fig. 1) (Miller et al., 2005), which predict increased rates of lineage divergence in the context of Sundaland (Fig. 2). This prediction is not supported here. Constant lineage divergence models of diversification are the most likely for all clades (Fig. S2; Table S3). In a model with a constant probability of divergence per lineage through time, lineage diversity increases linearly over time without levelling off toward the present (Table S2). Thus, the Pleistocene origin of a large proportion of lineages could be a consequence of constant diversification through time. This result might be surprising when considering the vast array of aquatic habitats occupied by Rasborinae in Sundaland: lowland vs. highland, fast vs. slow running waters, peat swamps, lakes, ponds (Brittan, 1972; Hubert et al., 2019; Kottelat et al., 1993; Kusuma et al., 2016; Sholihah et al., 2020). Besides, ecological transitions are scarce in Rasborinae. Clades I, II and III include small-size, forest-associated species while Clade IV include all the large, riverine and open-habitat *Rasbora* species (Kottelat et al., 1993; Liao et al., 2011; Sholihah et al., 2020). This pattern suggests that adaptive habitat-shift had limited impact on the pace of diversification in Rasborinae. Yet, Clade IV has the highest diversification rate of the four clades and is also the only clade that colonized Java. Altogether, the transition between forest and open habitat, which happened during the onset of Clade IV diversification, and higher divergence rate suggest an influence of ecological contingency on Rasborinae diversification. This group of riverine and opportunistic species was successful at colonizing open habitats, a previously non-colonized set of habitats for Rasborinae.

Despite the high diversity of mitochondrial lineages in Sundaland Rasborinae, we find no evidence of diversity equilibrium. The diversity-dependent diversification models received little support for all clades, suggesting intra-clade biotic interactions have little influence on either divergence opportunities or probabilities of maintenance of newly emerging lineages (Alonso, Etienne, & McKane, 2006; Hubert, Calcagno, Etienne, & Mouquet, 2015). If considering the ample distribution and abundance of Rasborinae in Sundaland and their presence in all aquatic habitats, this result is surprising. Rasborinae assemblages were likely

unsaturated during the diversification of the subfamily (Cornell, 1993). Diversity equilibrium and slowdown of diversification rates implies a shift in divergence/extinction equilibriums once assemblages are saturated (Hubert, Calcagno, et al., 2015; Kisel, McInnes, Toomey, & Orme, 2011; Phillimore & Price, 2008). Here, PCFs might be expected to have cyclically perturbed aquatic assemblages, and regularly disrupted a potential course toward equilibrium. Alternatively, Sundaland ecological carrying capacity might still be far from being reached, despite the staggering species richness of Sundaland ichthyofauna.

4.4 Robustness of the inferences and systematic implications

Our study first confirms the monophyly of the subfamily Rasborinae and its distinctiveness from the subfamilies Chedrinae, Danioninae and Esominae (Conway, Chen, & Mayden, 2008; Fang, Norin, Liao, Källersjö, & Kullander, 2009; Mayden et al., 2007; Rüber, Kottelat, Tan, Ng, & Britz, 2007; Tan & Armbruster, 2018; Tang et al., 2010). Our clade-specific phylogenetic reconstructions confirm the monophyly of the genera *Boraras*, *Pectenocypris* and *Trigonostigma*, but also support the paraphyly of the genus *Trigonopoma*, and the polyphyly of the genera *Brevibora* and *Rasbora*. The polyphyly of *Rasbora* is in line with previous molecular and morphology-based phylogenetic studies, which highlighted that *Rasbora* encompasses several lineages of distinct evolutionary origins (Liao et al., 2010, 2011; Lumbantobing, 2010; Sholihah et al., 2020; M. Tan & Armbruster, 2018; K. L. Tang et al., 2010). Yet, most *Rasbora* species groups initially described by Brittan (1954, 1972) are recovered. For instance, Clade III matches the delimitation of *R. einthovenii* species group and Clade II matches the boundaries of *R. trifasciata* species group (Brittan, 1972; Liao et al., 2010). We also support the elevation of *R. pauciperforatum* species group to the genus level (*Trigonopoma*) by Liao et al (2010). However, our results indicate *R. kalbarensis* should probably be considered a member of the genus *Trigonopoma*. Likewise, the genus *Boraras* described by Kottelat and Vidthayanon (Kottelat & Vidthayanon, 1993) is supported here. Most of the taxonomic conflicts are concentrated in Clade IV, in which *R. lateristriata* and *R. sumatrana* species groups from Brittan (1972), and the genus *Brevibora* from Liao et al (2011) are not monophyletic. The present study warrants further taxonomic works within Rasborinae and highlights the need of an in-depth revision of the genera *Rasbora* and *Brevibora*.

The estimated age of the Rasborinae mitochondrial MRCA, dated at ca. 46 Ma (41-53 Ma), is consistent with the 44 Myrs reported by Betancur-R et al. (2017) based on one mitochondrial and 20 nuclear genes. However, these results are not concordant with the ray-

finned fishes phylogeny of Rabosky et al. (2018) dating the MRCA of Rasborinae at 59 Ma, and the MRCA of Sundaland Rasborinae at 42 Ma. Several differences in the calibration methods and molecular markers may account for these differences. Phylogenetic reconstructions of Rabosky et al. (2018) are based on 20 nuclear markers. Discordances between mitochondrial and nuclear markers might be expected due to the maternal inheritance of mitochondrial genomes, with potential cryptic past introgression events and selective sweeps leading to conflicting phylogenetic reconstructions (Hubert, Torrico, Bonhomme, & Renno, 2007; Ríos, Bouza, & García, 2019; Wallis et al., 2017). Rabosky et al. (2018) and our phylogenetic reconstructions yielded concordant deep phylogenetic relationships, with lineages of Asia mainland (*Amblypharyngodon*, *Horadandia* and *Rasboroides*) placed in basal positions, and Sundaland lineages being equally assorted into the four clades I-IV. However, discordance between mitochondrial and nuclear markers due to past introgression cannot be discarded for shallower phylogenetic scales, hybridization being common in Cypriniformes (Atsumi, Nomoto, Machida, Ichimura, & Koizumi, 2018; Hopkins & Eisenhour, 2008; Tang, Liu, Yu, Liu, & Danley, 2012). Discrepancies in the age estimate of Rasborinae MRCA are most likely due to the use of alternative calibration methods. Rabosky et al. (Rabosky et al., 2018) fossil calibrations are scarce for Cypriniformes, with only three, distantly related from Rasborinae and ancient fossils (†*Amyzon aggregatum* for the North American Catostomidae dated between 49-122 Ma, †*Cyprinus maomingensis* for Eurasia Cyprini dated between 40-98 Ma, and †*Huashancyprinus robustispinus* for Eurasia *Cyprinus* dated between 23-73 Ma). The Rasborinae MRCA age estimate proposed by Rabosky et al. (2018) implies a markedly different biogeographic scenario, with Rasborinae diversification originating much before the isolation of Borneo. This ancient origin is not supported by recent phylogenetic reconstructions among the Cypriniformes genera *Nemacheilus* and *Pangio* in Southeast Asia, which MRCAs were recently dated at ca. 20 Ma according to phylogenetic reconstructions based on both mitochondrial and nuclear markers (Šlechtová et al., 2021). Along the same line, a recent reappraisal of previously published phylogenetic reconstructions support an early to late Miocene origin of several Southeast Asian fish taxa (Sholihah et al., 2021).

5. CONCLUSION

Pleistocene sea level changes had less impact on Rasborinae diversification than expected under the Paleoriver Hypothesis. Geographical patterns of divergence suggest that limited dispersal abilities and *in situ* diversification were predominant during the diversity

build-up of Rasborinae in Sundaland. Our phylogenetic and biogeographic reconstructions are in line with the timeframe of geological reconstructions in Southeast Asia. The Rasborinae followed the geological history of Sundaland and tightly matched the onset of isolation of Borneo. In particular, the North Sunda paleoriver is identified as a key aquatic system during the rise of the group and an important source of diversity for the neighbouring river systems. Our study also provides new lines of evidence about dispersal of freshwater organisms in Sundaland. Paleoriver watershed boundaries poorly explain geographical patterns of divergence and dispersal between islands. Surprisingly, our macroevolutionary inferences show no evidence of diversification slowdown and diversity ceiling, despite the exceptional levels of species richness of Sundaland ichthyofauna. This unexpected result further questions the mechanisms underlying these diversification trends. Several alternative scenarios may be invoked, including the impact of PCFs on disturbing speciation/extinction equilibriums. Our study warrants further research on the evolutionary mechanisms underlying diversification in such species-rich tropical systems.

Acknowledgements

The authors wish to thank Siti Nuramaliati Prijono, Bambang Sunarko, Witjaksono, Mohammad Irham, Marlina Adriyani, Ruliyana Susanti, Rosichon Ubaidillah, the late Renny K. Hadiaty, Hari Sutrisno and Cahyo Rahmadi at Research Centre for Biology (RCB-LIPI) in Indonesia; Edmond Dounias, Jean-Paul Toutain, Robert Arfi and Valérie Verdier from the ‘Institut de Recherche pour le Développement’; Joel Le Bail and Nicolas Gascoin at the French embassy in Jakarta for their continuous support. We also would like to thank Eleanor Adamson, Hendry Budianto, Tob Chann Aun, Pak Epang, Herman Ganatpathy, Renny K Hadiaty, Sébastien Lavoué, Michael Lo, Hendry Michael, Sopian Sauri, Joshua Siow, Heok Hui Tan, Elango Velautham, Norsham S. Yaakob, and Denis Yong for their help in the field. We are also particularly thankful to Sumanta at IRD Jakarta for his help during the field sampling in Indonesia. The present study and all associated methods were carried out in accordance with relevant guidelines and regulation of the Indonesian Ministry of Research and Technology (Indonesia), the Economic Planning Unit, Prime Minister’s Department (Malaysia), the Forest Department Sarawak (Malaysia), the Vietnam National Museum of Nature (Vietnam) and the Inland Fisheries Research and Development Institute (Cambodia). Field sampling in Indonesia was conducted according to the research permits 7/TKPIPA/FRP/SM/VII/2012, 68/EXT/SIP/FRP/SM/VIII/2013, 41/EXT/SIP/FRP/SM/VIII/2014, 361/SIP/FRP/E5/Dit.KI/IX/2015,

50/EXT/SIP/FRP/E5/Dit.KI/IX/2016, 45/EXT/SIP/FRP/E5/Dit.KI/VIII/2017, and
392/SIP/FRP/E5/Dit.KI/XI/2018 for Nicolas Hubert, and, 1/TKPIPA/FRP/ SM/I/2011 and
3/TKPIPA/FRP/SM/III/2012 for Lukas Rüber. The Fieldwork in Peninsular Malaysia and
Sarawak was conducted under permits issued by the Economic Planning Unit, Prime
Minister's Department, Malaysia (UPE 40/200/19/2417 and UPE 40/200/19/2534) and the
Forest Department Sarawak (NCCD.970.4.4[V]-43) and were obtained with the help of
Norsham S. Yaakob (Forest Research Institute Malaysia, Kepong, Kuala Lumpur, Malaysia).
Luong Van Hao and Pham Van Luc (Vietnam National Museum of Nature) helped with
arranging research permits in Vietnam and So Nam (Inland Fisheries Research and
Development Institute, IFRDI) helped with arranging research permits in Cambodia. All
experimental protocols were approved by the Indonesian Ministry of Research and
Technology (Indonesia), the Indonesian Institute of Sciences (Indonesia), the Forest
Department Sarawak (Malaysia), Economic Planning Unit of the Prime Minister's
Department (Malaysia), the Vietnam National Museum of Nature (Vietnam) and the Inland
Fisheries Research and Development Institute (Cambodia). This publication has ISEM
number 2021-062 SUD.

Data availability statement

DNA barcodes are available in the Barcode of Life Data System (dx.doi.org/110.5883/DS-BIFRA) and GenBank (see Table S1 for accession numbers). Mitogenomes are available in
GenBank (see Table S1 for accession numbers). Phylogenetic reconstructions and DNA
sequence alignments are available in Dryad (doi.org/10.5061/dryad.tb2rbp00g).

References

- Adamson, E. A. S., Hurwood, D. A., & Mather, P. B. (2010). A reappraisal of the evolution
of Asian snakehead fishes (Pisces, Channidae) using molecular data from multiple genes
and fossil calibration. *Molecular Phylogenetics and Evolution*, 56(2), 707–717.
<https://doi.org/10.1016/j.ympev.2010.03.027>
- Alonso, D., Etienne, R. S., & McKane, A. J. (2006). The merits of neutral theory. *Trends in
Ecology and Evolution*, 21(8), 451–457. <https://doi.org/10.1016/j.tree.2006.03.019>
- Atsumi, K., Nomoto, K., Machida, Y., Ichimura, M., & Koizumi, I. (2018). No reduction of
hatching rates among F1 hybrids of naturally hybridizing three Far Eastern daces, genus

- 781 Tribolodon (Cypriniformes, Cyprinidae). *Ichthyological Research*, 65(1), 165–167.
- 782 Barraclough, T. G., & Nee, S. (2001). Phylogenetics and Speciation. *Trends in Ecology and*
783 *Evolution*, 16(7), 391–399.
- 784 Beck, S. V., Carvalho, G. R., Barlow, A., Ruber, L., Hui Tan, H., Nugroho, E., ... de Bruyn,
785 M. (2017). Plio-Pleistocene phylogeography of the Southeast Asian Blue Panchax
786 killifish, *Aplocheilichthys panchax*. *PLoS One*, 12(7), e0179557.
787 <https://doi.org/10.1371/journal.pone.0179557>
- 788 Bermingham, E., McCafferty, S., & Martin, A. P. (1997). Fish biogeography and molecular
789 clocks: perspectives from the Panamanian isthmus. In T. D. Kocher & C. A. Stepien
790 (Eds.), *Molecular systematics of fishes* (pp. 113–128). San Diego: CA Academic Press.
- 791 Bernatchez, L., & Wilson, C. (1998). Comparative phylogography of Neartic and Palearctic
792 fishes. *Molecular Ecology*, 7, 431–452.
- 793 Betancur-R, R., Wiley, E. O., Arratia, G., Acero, A., Bailly, N., Miya, M., ... Ortí, G. (2017).
794 Phylogenetic classification of bony fishes. *BMC Evolutionary Biology*, 17(1), 162.
795 <https://doi.org/10.1186/s12862-017-0958-3>
- 796 Bird, M. I., Taylor, D., & Hunt, C. (2005). Palaeoenvironments of insular Southeast Asia
797 during the Last Pleistocene. Glacial Period: a savanna corridor in Sundaland? .
798 *Quaternary Science Reviews*, 24, 2228–2242.
- 799 Blaxter, M., Mann, J., Chapman, T., Thomas, F., Whitton, C., Floyd, R., & Abebe, E. (2005).
800 Defining operational taxonomic units using DNA barcode data. *Philosophical*
801 *Transactions of the Royal Society B: Biological Sciences*, 360(1462), 1935–1943.
- 802 Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C. H., Xie, D., ... Drummond, A. J.
803 (2014). BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. *PLoS*
804 *Computational Biology*, 10(4), 1–6. <https://doi.org/10.1371/journal.pcbi.1003537>
- 805 Brittan, M. R. (1954). A revision of the Indo-Malayan frash-water fish genus *Rasbora*.
806 *Monogr. Inst. Sci. Tech. Manila*, 3, 3–pls.
- 807 Brittan, M. R. (1972). *Rasbora: A revision of the Indo-Malayan freshwater fish genus*

808 *Rasbora*. Hongkong: TFH Publications.

809 Brown, R. M., Siler, C. D., Oliveros, C. H., Esselstyn, J. A., Diesmos, A. C., Hosner, P. A.,
810 ... Alcala, A. C. (2013). Evolutionary Processes of Diversification in a Model Island
811 Archipelago. *Annual Review of Ecology, Evolution, and Systematics*, 44(1), 411–435.
812 <https://doi.org/10.1146/annurev-ecolsys-110411-160323>

813 Cannon, C. H., Morley, R. J., & Bush, M. B. (2009). The current refugial rainforests of
814 Sundaland are unrepresentative of their biogeographic past and highly vulnerable to
815 disturbance. *Proceedings of the National Academy of Sciences, USA*, 106(27), 11188–
816 11193.

817 Chang, J., Rabosky, D. L., Smith, S. A., & Alfaro, M. E. (2019). An r package and online
818 resource for macroevolutionary studies using the ray-finned fish tree of life. *Methods in*
819 *Ecology and Evolution*, 10(7), 1118–1124.

820 Clark, J. R., Ree, R. H., Alfaro, M. E., King, M. G., Wagner, W. L., & Roalson, E. H. (2008).
821 A comparative study in ancestral range reconstruction methods: retracing the uncertain
822 histories of insular lineages. *Systematic Biology*, 57(5), 693–707.

823 Condamine, F. L., Rolland, J., & Morlon, H. (2013). Macroevo-lutionary perspectives to
824 environmental change. *Ecology Letters*, 16(SUPPL.1), 72–85.
825 <https://doi.org/10.1111/ele.12062>

826 Condamine, F. L., Rolland, J., & Morlon, H. (2019). Assessing the causes of diversification
827 slowdowns: temperature-dependent and diversity-dependent models receive equivalent
828 support. *Ecology Letters*, 22(11), 1900–1912. <https://doi.org/10.1111/ele.13382>

829 Condamine, F. L., Sperling, F. A. H., Wahlberg, N., Rasplus, J., & Kergoat, G. J. (2012).
830 What causes latitudinal gradients in species diversity? Evolutionary processes and
831 ecological constraints on swallowtail biodiversity. *Ecology Letters*, 15(3), 267–277.

832 Condamine, F. L., Toussaint, E. F. A., Clamens, A. L., Genson, G., Sperling, F. A. H., &
833 Kergoat, G. J. (2015). Deciphering the evolution of birdwing butterflies 150 years after
834 Alfred Russel Wallace. *Scientific Reports*, 5, 1–11. <https://doi.org/10.1038/srep11860>

- 835 Condamine, F. L., Toussaint, E. F. A., Cotton, A. M., Genson, G. S., Sperling, F. A. H., &
836 Kergoat, G. J. (2013). Fine-scale biogeographical and temporal diversification
837 processes of peacock swallowtails (*Papilio* subgenus *Achillides*) in the Indo-Australian
838 Archipelago. *Cladistics*, 29(1), 88–111.
- 839 Conway, K. W., Chen, W.-J., & Mayden, R. L. (2008). The “Celestial Pearl danio” is a
840 miniature *Danio* (s.s.) (Ostariophysi: Cyprinidae): evidence from morphology and
841 molecules. *Zootaxa*, 1686, 1–28.
- 842 Cornell, H. V. (1993). Unsaturated patterns in species assemblages: the role of regional
843 processes in setting local species richness. In R. E. Ricklefs Schluter, D (Ed.), *Species*
844 *diversity in ecological communities: historical and geographical perspectives* (pp. 243–
845 252). Chicago: University of Chicago Press.
- 846 Cowie, R. H., & Holland, B. S. (2006). Dispersal is fundamental to biogeography and the
847 evolution of biodiversity on oceanic islands. *Journal of Biogeography*, 33(2), 193–198.
848 <https://doi.org/10.1111/j.1365-2699.2005.01383.x>
- 849 Dahrudin, H., Hutama, A., Busson, F., Sauri, S., Hanner, R., Keith, P., ... Hubert, N. (2017).
850 Revisiting the ichthyodiversity of Java and Bali through DNA barcodes: taxonomic
851 coverage, identification accuracy, cryptic diversity and identification of exotic species.
852 *Molecular Ecology Resources*, 17(2), 288–299. [https://doi.org/10.1111/1755-](https://doi.org/10.1111/1755-0998.12528)
853 [0998.12528](https://doi.org/10.1111/1755-0998.12528)
- 854 de Bruyn, M., Rüber, L., Nylinder, S., Stelbrink, B., Lovejoy, N. R., Lavoué, S., ... Carvalho,
855 G. R. (2013). Paleo-drainage basin connectivity predicts evolutionary relationships
856 across three Southeast asian biodiversity hotspots. *Systematic Biology*, 62(3), 398–410.
- 857 de Bruyn, M., Stelbrink, B., Morley, R. J., Hall, R., Carvalho, G. R., Cannon, C. H., ... von
858 Rintelen, T. (2014). Borneo and Indochina are Major Evolutionary Hotspots for
859 Southeast Asian Biodiversity. *Syst Biol*, 63(6), 23. <https://doi.org/10.5061/dryad.67s40>
- 860 Dodson, J. J., Colombani, F., & Ng, P. K. L. (1995). Phylogeographic structure in
861 mitochondrial DNA of a South-east Asian freshwater fish, *Hemibagrus nemurus*
862 (Sluroidei; Bagridae) and Pleistocene sea-level changes on the Sunda shelf. *Molecular*
863 *Ecology*, 4, 331–346.

- 864 Dodsworth, S. (2015). Genome skimming for next-generation biodiversity analysis. *Trends in*
865 *Plant Science*, 20(9), 525–527.
- 866 Drummond, A., Suchard, M., Xie, D., & Rambault, A. (2012). Bayesian phylogenetics with
867 BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, 29(1969–1973).
- 868 Durand, J., Persat, H., & Bouvet, Y. (1999). Phylogeography and postglacial dispersion of the
869 chub (*Leuciscus cephalus*) in Europe. *Molecular Ecology*, 8, 989–997.
- 870 Eschmeyer, W. N., Fricke, R., & van der Laan, R. (2018). Catalog of fishes electronic version
871 (C. A. of Sciences, Ed.).
- 872 Esselstyn, J. A., & Brown, R. M. (2009). The role of repeated sea-level fluctuations in the
873 generation of shrew (Soricidae: Crocidura) diversity in the Philippine Archipelago.
874 *Molecular Phylogenetics and Evolution*, 53(1), 171–181.
875 <https://doi.org/10.1016/j.ympev.2009.05.034>
- 876 Esselstyn, J. A., Timm, R. M., & Brown, R. M. (2009). Do geological or climatic processes
877 drive speciation in dynamic archipelagos? the tempo and mode of diversification in
878 southeast asian shrews. *Evolution*, 63(10), 2595–2610. [https://doi.org/10.1111/j.1558-](https://doi.org/10.1111/j.1558-5646.2009.00743.x)
879 [5646.2009.00743.x](https://doi.org/10.1111/j.1558-5646.2009.00743.x)
- 880 Etienne, R. S., Haegeman, B., Stadler, T., Aze, T., Pearson, P. N., Purvis, A., & Phillimore,
881 A. B. (2012). Diversity-dependence brings molecular phylogenies closer to agreement
882 with the fossil record. *Proceedings of the Royal Society B: Biological Sciences*,
883 279(1732), 1300–1309. <https://doi.org/10.1098/rspb.2011.1439>
- 884 Fang, F., Norin, M., Liao, T. Y., Källersjö, M., & Kullander, S. O. (2009). Molecular
885 phylogenetic interrelationships of the south Asian cyprinid genera Danio, Devario and
886 Microrasbora (Teleostei, Cyprinidae, Danioninae). *Zoologica Scripta*, 38, 237–256.
- 887 Froese, R., & Pauly, D. (2020). Fishbase. Retrieved from Worldwide web electronic
888 publication, www.fishbase.org, version (01/2020) website: <http://www.fishbase.org>
- 889 Fujiwara, T., & Barraclough, T. G. (2013). Delimiting species using single-locus data and the
890 generalized mixed yule coalescent approach: a revised method and evaluation on

891 simulated data sets. *Systematic Biology*, 62(5), 707–724.

892 Garg, T., Hamilton, S. E., Hochard, J. P., Kresch, E. P., & Talbot, J. (2018). (Not so) gently
893 down the stream: River pollution and health in Indonesia. *Journal of Environmental*
894 *Economics and Management*, 92, 35–53.

895 Gaveau, D. L. A., Sloan, S., Molidena, E., Yaen, H., Sheil, D., Abram, N. K., ... Wielaard, N.
896 (2014). Four decades of forest persistence, clearance and logging on Borneo. *PloS One*,
897 9(7).

898 Gorog, A. J., Sinaga, M. H., & Engstrom, M. D. (2004). Vicariance or dispersal? Historical
899 biogeography of three Sunda shelf marine rodents (*Maxomys surifer*, *Leopoldamys*
900 *sabanus* and *Maxomys whiteheadi*). *Biological Journal of the Linnean Society*, 81(1),
901 91–109. <https://doi.org/10.1111/j.1095-8312.2004.00281.x>

902 Hall, R. (2009). Southeast Asia's changing palaeogeography. *Blumea: Journal of Plant*
903 *Taxonomy and Plant Geography*, 54(1–3), 148–161.
904 <https://doi.org/10.3767/000651909X475941>

905 Hall, R. (2013). The palaeogeography of Sundaland and Wallacea since the Late Jurassic.
906 *Journal of Limnology*, 72(S2), 1–17. <https://doi.org/10.4081/jlimnol.2013.s2.e1>

907 Hall, R., Cottam, M. A., & Wilson, M. E. J. (2011). The SE Asian gateway: history and
908 tectonics of the Australia–Asia collision. *Geological Society, London, Special*
909 *Publications*, 355(1), 1–6. <https://doi.org/10.1144/SP355.1>

910 Hardman, M., & Lundberg, J. G. (2006). Molecular phylogeny and chronology of
911 diversification for “phractocephaline” catfishes (Siluriformes: Pimelodidae) based on
912 mitochondrial DNA and DNA recombination activating gene 2 sequences. *Molecular*
913 *Phylogenetics and Evolution*, 40, 410–418.

914 Heaney, L. R. (1992). A synopsis of climatic and vegetational change in Southeast Asia.
915 *Tropical Forests and Climate*, 53–61. https://doi.org/10.1007/978-94-017-3608-4_6

916 Heled, J., & Drummond, A. J. (2010). Bayesian Inference of Species Trees from Multilocus
917 Data. *Molecular Biology and Evolution*, 27(3), 570–580.

918 <https://doi.org/10.1093/molbev/msp274>

919 Hendriks, K. P., Alciatore, G., Schilthuizen, M., & Etienne, R. S. (2019). Phylogeography of
920 Bornean land snails suggests long-distance dispersal as a cause of endemism. *Journal*
921 *of Biogeography*.

922 Ho, S. Y. W., & Larson, G. (2006). Molecular clocks: when times are a-changin'. *TRENDS in*
923 *Genetics*, 22(2), 79–83.

924 Hoffmann, M., Hilton-Taylor, C., Angulo, A., Böhm, M., Brooks, T. M., Butchart, S. H. M.,
925 ... Stuart, S. N. (2010). The impact of conservation on the status of the world's
926 vertebrates. *Science*, 330(6010), 1503–1509. <https://doi.org/10.1126/science.1194442>

927 Hopkins, R. L., & Eisenhour, D. J. (2008). Hybridization of *Lythrurus fasciolaris* and
928 *Lythrurus umbratilis* (Cypriniformes: Cyprinidae) in the Ohio River basin. *Copeia*,
929 2008(1), 162–171.

930 Hubert, N., Calcagno, V., Etienne, R. S., & Mouquet, N. (2015). Metacommunity speciation
931 models and their implication for diversification theory. *Ecology Letters*, 18(8), 864–881.

932 Hubert, N., Calcagno, V., Etienne, R. S., & Mouquet, N. (2015). Metacommunity speciation
933 models and their implications for diversification theory. *Ecology Letters*, Vol. 18, pp.
934 864–881. <https://doi.org/10.1111/ele.12458>

935 Hubert, N., Duponchelle, F., Nuñez, J., Garcia-Davila, C., Paugy, D., & Renno, J. F. (2007).
936 Phylogeography of the piranha genera *Serrasalmus* and *Pygocentrus*: Implications for
937 the diversification of the Neotropical ichthyofauna. *Molecular Ecology*, 16(10), 2115–
938 2136. <https://doi.org/10.1111/j.1365-294X.2007.03267.x>

939 Hubert, N., Kadarusman, Wibowo, A., Busson, F., Caruso, D., Sulandari, S., ... Hadiaty, R.
940 K. (2015). DNA barcoding Indonesian freshwater fishes: challenges and prospects. *DNA*
941 *Barcodes*, 3, 144–169.

942 Hubert, N., Lumbantobing, D., Sholihah, A., Dahruddin, H., Delrieu-Trottin, E., Busson, F.,
943 ... Keith, P. (2019). Revisiting species boundaries and distribution ranges of
944 *Nemacheilus* spp. (Cypriniformes: Nemacheilidae) and *Rasbora* spp. (Cypriniformes:

945 Cyprinidae) in Java, Bali and Lombok through DNA barcodes: implications for
 946 conservation in a biodiversity hotspot. *Conservation Genetics*, 20(3), 517–529.
 947 <https://doi.org/10.1007/s10592-019-01152-w>

948 Hubert, N., & Renno, J. F. (2006). Historical Biogeography of South American Freshwater
 949 fishes. *Journal of Biogeography*, 33, 1414–1436.

950 Hubert, N., Torrico, J. P., Bonhomme, F., & Renno, J. F. (2007). Species polyphyly and
 951 mtDNA introgression among three *Serrasalmus* sister-species. *Molecular Phylogenetics*
 952 *and Evolution*, 46, 375–381.

953 Hui, T. H., & Chian, L. K. (2020). *Trigonostigma truncata*, a new species of harlequin
 954 rasbora from Malay Peninsula (Teleostei: Danionidae). *RAFFLES BULLETIN OF*
 955 *ZOOLOGY*, 7600(May). <https://doi.org/10.26107/RBZ-2020-0058>

956 Husson, L., Boucher, F. C., Sarr, A. C., Sepulchre, P., & Cahyarini, S. Y. (2020). Evidence of
 957 Sundaland's subsidence requires revisiting its biogeography. *Journal of Biogeography*,
 958 Vol. 47, pp. 843–853. <https://doi.org/10.1111/jbi.13762>

959 Hutama, A., Dahrudin, H., Busson, F., Sauri, S., Keith, P., Hadiaty, R. K., ... Hubert, N.
 960 (2017). Identifying spatially concordant evolutionary significant units across multiple
 961 species through DNA barcodes: Application to the conservation genetics of the
 962 freshwater fishes of Java and Bali. *Global Ecology and Conservation*, 12, 170–187.
 963 <https://doi.org/10.1016/j.gecco.2017.11.005>

964 Imai, N., Furukawa, T., Tsujino, R., Kitamura, S., & Yumoto, T. (2018). Factors affecting
 965 forest area change in Southeast Asia during 1980-2010. *PloS One*, 13(5), e0197391.

966 Ivanova, N. V, Zemlak, T. S., Hanner, R. H., & Hébert, P. D. N. (2007). Universal primers
 967 cocktails for fish DNA barcoding. *Molecular Ecology Notes*, 7(4), 544–548.

968 Iwasaki, W., Fukunaga, T., Isagozawa, R., Yamada, K., Maeda, Y., Satoh, T. P., ... Miya, M.
 969 (2013). MitoFish and MitoAnnotator: a mitochondrial genome database of fish with an
 970 accurate and automatic annotation pipeline. *Molecular Biology and Evolution*, 30(11),
 971 2531–2540.

- 972 Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., & Jermin, L. S.
973 (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature*
974 *Methods*, 14(6), 587–589.
- 975 Kisel, Y., McInnes, L., Toomey, N. H., & Orme, C. D. L. (2011). How diversification rates
976 and diversity limits combine to create large-scale species-area relationships.
977 *Philosophical Transactions of the Royal Society B-Biological Sciences*, 366(1577),
978 2514–2525. <https://doi.org/10.1098/rstb.2011.0022>
- 979 Kottelat, M. (2013). The fishes of the inland waters of Southeast Asia: a catalog and core
980 bibliography of the fishes known to occur in freshwaters, mangroves and estuaries. *The*
981 *Raffles Bulletin of Zoology, Supplement*, 1–663.
- 982 Kottelat, M., & Vidthayanon, C. (1993). Boraras micros, a new genus and species of minute
983 freshwater fish from Thailand (Teleostei: Cyprinidae). *Ichthyol. Explor. Freshwaters*, 4,
984 161–176.
- 985 Kottelat, M., Whitten, A. J., Kartikasari, N., & Wirjoatmodjo, S. (1993). *Freshwater Fishes*
986 *of Western Indonesia and Sulawesi*. <https://doi.org/10.2307/1447208>
- 987 Kottelat, M., Whitten, A. J., Kartikasari, S. R., & Wirjoatmodjo, S. (1993). *Freshwater fishes*
988 *of western indonesia and sulawesi*. Singapore: Periplus editions.
- 989 Kottelat, M., & Witte, K.-E. (1999). Two new species of Microrasbora from Thailand and
990 Myanmar, with two new generic names for small Southeast Asian cyprinid fishes
991 (Teleostei: Cyprinidae). *Journal of South Asian Natural History*, 4(1), 49–56.
- 992 Kusuma, W. E., Ratmuangkhwang, S., & Kumazawa, Y. (2016). Molecular phylogeny and
993 historical biogeography of the Indonesian freshwater fish Rasbora lateristriata species
994 complex (Actinopterygii: Cyprinidae): Cryptic species and west-to-east divergences.
995 *Molecular Phylogenetics and Evolution*, 105, 212–223.
996 <https://doi.org/10.1016/j.ympev.2016.08.014>
- 997 Laumonier, Y., Uryu, Y., Stüwe, M., Budiman, A., Setiabudi, B., & Hadian, O. (2010). Eco-
998 floristic sectors and deforestation threats in Sumatra: identifying new conservation area
999 network priorities for ecosystem-based land use planning. *Biodiversity and*

- 1000 *Conservation*, 19(4), 1153–1174.
- 1001 Liao, T. Y., Kullander, S. O., & Fang, F. (2010). Phylogenetic analysis of the genus *Rasbora*
1002 (Teleostei: Cyprinidae). *Zoologica Scripta*, 39(2), 155–176.
- 1003 Liao, T. Y., Kullander, S. O., & Fang, F. (2011). Phylogenetic position of rasborin cyprinids
1004 and monophyly of major lineages among the Danioninae, based on morphological
1005 characters (Cypriniformes: Cyprinidae). *Journal of Zoological Systematics and*
1006 *Evolutionary Research*, 49(3), 224–232. [https://doi.org/10.1111/j.1439-](https://doi.org/10.1111/j.1439-0469.2011.00621.x)
1007 0469.2011.00621.x
- 1008 Lim, N. K. M., Tay, Y. C., Srivathsan, A., Tan, J. W. T., Kwik, J. T. B., Baloglu, B., ... Yeo,
1009 D. C. J. (2016). Next-generation freshwater bioassessment: eDNA metabarcoding with a
1010 conserved metazoan primer reveals species-rich and reservoir-specific communities.
1011 *Royal Society Open Science*, 3(11), 160635.
- 1012 Logue, J. B., Mouquet, N., Hannes, P., Hillebrand, H., & group, T. metacommunity working.
1013 (2011). Empirical approaches to metacommunities: a review and comparison with
1014 theory. *Trends in Ecology and Evolution*, 26(9), 482–491.
- 1015 Lohman, K., De Bruyn, M., Page, T., Von Rintelen, K., Hall, R., Ng, P. K. L., ... Von
1016 Rintelen, T. (2011). Biogeography of the Indo-Australian archipelago. *Annual Review of*
1017 *Ecology, Evolution and Systematics*, 42, 205–226.
- 1018 Lumbantobing, D. N. (2010). Analisis filogenetik genus *Rasbora* (Teleostei: Cyprinidae)
1019 berdasarkan karakter morfologis. *Jurnal Iktiologi Indonesia*, 10(2), 185–189.
- 1020 Matzke, N. J. (2013). Probabilistic historical biogeography: new models for founder-event
1021 speciation, imperfect detection, and fossils allow improved accuracy and model-testing.
1022 *Frontiers of Biogeography*, 5(4), 242–248. <https://doi.org/10.21425/f55419694>
- 1023 Matzke, N. J. (2014). Model selection in historical biogeography reveals that founder-event
1024 speciation is a crucial process in island clades. *Systematic Biology*, 63(6), 951–970.
1025 <https://doi.org/10.1093/sysbio/syu056>
- 1026 Mayden, R. L., Tang, K. L., Conway, K. W., Freyhof, J., Chamberlain, S., Haskins, M., ...

1027 He, S. (2007). Phylogenetic relationships of *Danio* within the order Cypriniformes: a
 1028 framework for comparative and evolutionary studies of a model species. *Journal of*
 1029 *Experimental Zoology*, 308B, 642–654.

1030 Meyer, M., & Kircher, M. (2010). Illumina sequencing library preparation for highly
 1031 multiplexed target capture and sequencing. *Cold Spring Harbor Protocols*, 2010(6),
 1032 pdb-prot5448.

1033 Miller, K. G., Kominz, M. A., Browning, J. V, Wright, J. D., Mountain, G. S., Katz, M. E.,
 1034 ... Pekar, S. F. (2005). The phanerozoic record of global sea-level change. *Science*, Vol.
 1035 310, pp. 1293–1298. <https://doi.org/10.1126/science.1116412>

1036 Miller, M. A., Pfeiffer, W., & Schwartz, T. (2010). Creating the CIPRES Science Gateway
 1037 for inference of large phylogenetic trees. *2010 Gateway Computing Environments*
 1038 *Workshop, GCE 2010*. <https://doi.org/10.1109/GCE.2010.5676129>

1039 Mittelbach, G. G., Schemske, D. W., Cornell, H. V, Allen, A. P., Brown, J. M., Bush, M. B.,
 1040 ... Turelli, M. (2007). Evolution and the latitudinal diversity gradient: speciation,
 1041 extinction and biogeography. *Ecology Letters*, 10, 315–331.

1042 Mittermeier, R. A., Turner, W. R., Larsen, F. W., Brooks, T. M., & Gascon, C. (2011).
 1043 Global biodiversity conservation: the critical role of hotspots. In *Biodiversity hotspots*
 1044 (pp. 3–22). Springer.

1045 Morlon, H., Lewitus, E., Condamine, F. L., Manceau, M., Clavel, J., & Drury, J. (2016).
 1046 RPANDA: an R package for macroevolutionary analyses on phylogenetic trees.
 1047 *Methods in Ecology and Evolution*, 7(5), 589–597.

1048 Mulligan, M., van Soesbergen, A., & Sáenz, L. (2020). GOODD, a global dataset of more
 1049 than 38,000 georeferenced dams. *Scientific Data*, 7(1), 1–8.

1050 Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A. B., & Kent, F. (2000).
 1051 Biodiversity hotspots for conservation priorities. *Nature*, 403, 853–858.

1052 Nguyen, L.-T., Schmidt, H. A., Von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: a fast
 1053 and effective stochastic algorithm for estimating maximum-likelihood phylogenies.

1054 *Molecular Biology and Evolution*, 32(1), 268–274.

1055 Nguyen, T. T. T., Na-Nakorn, U., Sukmanomon, S., & ZiMing, C. (2008). A study on
 1056 phylogeny and biogeography of mahseer species (Pisces: Cyprinidae) using sequences
 1057 of three mitochondrial DNA gene regions. *Molecular Phylogenetics and Evolution*.
 1058 <https://doi.org/10.1016/j.ympev.2008.01.006>

1059 Nores, M. (1999). An alternative hypothesis for the origin of Amazonian bird diversity.
 1060 *Journal of Biogeography*, 26(3), 475–485. [https://doi.org/10.1046/j.1365-](https://doi.org/10.1046/j.1365-2699.1999.t01-1-00311.x)
 1061 [2699.1999.t01-1-00311.x](https://doi.org/10.1046/j.1365-2699.1999.t01-1-00311.x)

1062 Nurul Farhana, S., Muchlisin, Z. A., Duong, T. Y., Tanyaros, S., Page, L. M., Zhao, Y., ...
 1063 Siti Azizah, M. N. (2018). Exploring hidden diversity in Southeast Asia's *Dermogenys*
 1064 spp. (Belontiiformes: Zenarchopteridae) through DNA barcoding. *Scientific Reports*, 8(1),
 1065 10787. <https://doi.org/10.1038/s41598-018-29049-7>

1066 Ogilvie, H. A., Bouckaert, R. R., & Drummond, A. J. (2017). StarBEAST2 brings faster
 1067 species tree inference and accurate estimates of substitution rates. *Molecular Biology*
 1068 *and Evolution*, 34(8), 2101–2114. <https://doi.org/10.1093/molbev/msx126>

1069 Okonechnikov, K., Golosova, O., Fursov, M., Varlamov, A., Vaskin, Y., Efremov, I., ...
 1070 Tleukenov, T. (2012). Unipro UGENE: A unified bioinformatics toolkit. *Bioinformatics*,
 1071 28(8), 1166–1167. <https://doi.org/10.1093/bioinformatics/bts091>

1072 Orti, G. (1997). Radiation of Characiforms Fishes: Evidence from Mitochondrial and Nuclear
 1073 DNA sequences. In *Molecular Systematics of Fishes*. Academic Press.

1074 Papadopoulou, A., & Knowles, L. L. (2015a). Genomic tests of the species-pump hypothesis:
 1075 Recent island connectivity cycles drive population divergence but not speciation in
 1076 Caribbean crickets across the Virgin Islands. *Evolution*, 69(6), 1501–1517.
 1077 <https://doi.org/10.1111/evo.12667>

1078 Papadopoulou, A., & Knowles, L. L. (2015b). Species-specific responses to island
 1079 connectivity cycles: Refined models for testing phylogeographic concordance across a
 1080 Mediterranean Pleistocene Aggregate Island Complex. *Molecular Ecology*, 24(16),
 1081 4252–4268. <https://doi.org/10.1111/mec.13305>

1082 Paradis, E., & Schliep, K. (2019). ape 5.0: an environment for modern phylogenetics and
 1083 evolutionary analyses in R. *Bioinformatics*, 35(3), 526–528.

1084 Patel, S., Weckstein, J. D., Patane, J. S., Bates, J. M., & Aleixo, A. (2011). Temporal and
 1085 spatial diversification of *Pteroglossus aracaris* (AVES: Ramphastidae) in the neotropics:
 1086 constant rate of diversification does not support an increase in radiation during the
 1087 Pleistocene. *Mol Phylogenet Evol*, 58(1), 105–115.
 1088 <https://doi.org/10.1016/j.ympev.2010.10.016>

1089 Phillimore, A. B., & Price, T. D. (2008). Density-dependent cladogenesis in birds. *PLoS*
 1090 *Biology*, 6(3), e71. <https://doi.org/10.1371/journal.pbio.0060071>

1091 Pouyaud, L., Sudarto, & Paradis, E. (2009). The phylogenetic structure of habitat shift and
 1092 morphological convergence in Asian *Clarias* (Teleostei, Siluriformes: Clariidae).
 1093 *Journal of Zoological Systematics and Evolutionary Research*, 47(4), 344–356.
 1094 <https://doi.org/10.1111/j.1439-0469.2008.00507.x>

1095 Puillandre, N., Lambert, A., Brouillet, S., & Achaz, G. (2012). ABGD, Automatic Barcode
 1096 Gap Discovery for primary species delimitation. *Molecular Ecology*, 21, 1864–1877.

1097 Rabosky, D. L., Chang, J., Title, P. O., Cowman, P. F., Sallan, L., Friedman, M., ... Coll, M.
 1098 (2018). An inverse latitudinal gradient in speciation rate for marine fishes. *Nature*,
 1099 559(7714), 392.

1100 Ratnasingham, S., & Hebert, P. D. N. (2007). BOLD: The Barcode of Life Data System
 1101 (www.barcodinglife.org). *Molecular Ecology Notes*, 7(3), 355–364. <https://doi.org/DOI>
 1102 10.1111/j.1471-8286.2006.01678.x

1103 Ratnasingham, S., & Hebert, P. D. N. (2013). A DNA-Based Registry for All Animal
 1104 Species: The Barcode Index Number (BIN) System. *PLoS ONE*, 8(7).
 1105 <https://doi.org/10.1371/journal.pone.0066213>

1106 Ree, R. H., & Sanmartín, I. (2018). Conceptual and statistical problems with the DEC+ J
 1107 model of founder-event speciation and its comparison with DEC via model selection.
 1108 *Journal of Biogeography*, 45(4), 741–749.

- 1109 Ríos, N., Bouza, C., & García, G. (2019). Past hybridisation and introgression erased traces
1110 of mitochondrial lineages evolution in the Neotropical silver catfish *Rhamdia quelen*
1111 (Siluriformes: Heptapteridae). *Hydrobiologia*, 830(1), 161–177.
- 1112 Ritchie, A. M., Lo, N., & Ho, S. Y. W. (2017). The impact of the tree prior on molecular
1113 dating of data sets containing a mixture of inter-and intraspecies sampling. *Systematic*
1114 *Biology*, 66(3), 413–425.
- 1115 Rüber, L., Kottelat, M., Tan, H., Ng, P., & Britz, R. (2007). Evolution of miniaturization and
1116 the phylogenetic position of *Paedocypris*, comprising the world's smallest vertebrate.
1117 *BMC Evolutionary Biology*, 7(38).
- 1118 Saitoh, K., Sado, T., Mayden, R., Hanzawa, N., Nakamura, K., Nishida, M., & Miya, M.
1119 (2006). Mitogenomic evolution and interrelationships of the cypriniformes
1120 (Actinopterygii: Ostariophysi): the first evidence toward resolution of higher level
1121 relationships of the world's largest freshwater fish clade based on 59 whole mitogenome
1122 sequences. *Journal of Molecular Evolution*, 63, 826–841.
- 1123 Sarr, A., Sepulchre, P., & Husson, L. (2019). Impact of the Sunda Shelf on the Climate of the
1124 Maritime Continent. *Journal of Geophysical Research: Atmospheres*, 124(5), 2574–
1125 2588.
- 1126 Sathiamurthy, E., & Voris, K. H. (2006). Maps of Holocene Sea Level Transgression and
1127 Submerged Lakes on the Sunda Shelf. In *The Natural History Journal of Chulalongkorn*
1128 *University, Supplement* (Vol. 2).
- 1129 Sholihah, A., Delrieu-Trottin, E., Condamine, F. L., Wowor, D., Rüber, L., Pouyaud, L., ...
1130 Hubert, N. (2021). Impact of Pleistocene Eustatic Fluctuations on Evolutionary
1131 Dynamics in Southeast Asian Biodiversity Hotspots. *Systematic Biology*.
- 1132 Sholihah, A., Delrieu-Trottin, E., Sukmono, T., Dahrudin, H., Risdawati, R., Elvyra, R., ...
1133 Hubert, N. (2020). Disentangling the taxonomy of the subfamily Rasborinae
1134 (Cypriniformes, Danionidae) in Sundaland using DNA barcodes. *Scientific Reports*.
- 1135 Šlechtová, V., Musilova, Z., Hui Tan, H., Kottelat, M., & Bohlen, J. (2021). One northward,
1136 one southward: contrasting biogeographical history in two benthic freshwater fish

- 1137 genera across Southeast Asia (Teleostei: Cobitoidea: Nemacheilus, Pangio). *Molecular*
 1138 *Phylogenetics and Evolution*, 107139.
 1139 <https://doi.org/https://doi.org/10.1016/j.ympev.2021.107139>
- 1140 Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis
 1141 of large phylogenies. *Bioinformatics*, 30(9), 1312–1313.
- 1142 Stout, C. C., Tan, M., Lemmon, A. R., Lemmon, E. M., & Armbruster, J. W. (2016).
 1143 Resolving Cypriniformes relationships using an anchored enrichment approach. *BMC*
 1144 *Evolutionary Biology*, 16(1), 244.
- 1145 Straub, S. C. K., Parks, M., Weitemier, K., Fishbein, M., Cronn, R. C., & Liston, A. (2012).
 1146 Navigating the tip of the genomic iceberg: Next-generation sequencing for plant
 1147 systematics. *American Journal of Botany*, 99(2), 349–364.
- 1148 Swofford, D. L. (2001). Version 4.0 b10. *PAUP^{*}. Phylogenetic Analysis Using Parsimony*
 1149 *(^{*} and Other Methods)*.
- 1150 Tan, H. H., & Lim, K. K. P. (2013). Three new species of freshwater halfbeaks (Teleostei:
 1151 Zenarchopteridae: Hemirhamphodon) from Borneo. *Raffles Bulletin of Zoology*, 61(2),
 1152 735–747.
- 1153 Tan, M., & Armbruster, J. W. (2018). Phylogenetic classification of extant genera of fishes of
 1154 the order Cypriniformes (Teleostei: Ostariophysi). *Zootaxa*, 4476(1), 6–39.
- 1155 Tan, M. P., Jamsari, A. F. J., & Siti Azizah, M. N. (2012). Phylogeographic pattern of the
 1156 striped snakehead, *channa striata* in Sundaland: ancient river connectivity, geographical
 1157 and anthropogenic signatures. *PLoS ONE*, 7(12), e52089.
- 1158 Tang, K. L., Agnew, M. K., Hirt, M. V., Sado, T., Schneider, L. M., Freyhof, J., ... Mayden,
 1159 R. L. (2010). Systematics of the subfamily Danioninae (Teleostei: Cypriniformes:
 1160 Cyprinidae). *Molecular Phylogenetics and Evolution*, 57(1), 189–214.
 1161 <https://doi.org/10.1016/j.ympev.2010.05.021>
- 1162 Tang, Q., Liu, S., Yu, D., Liu, H., & Danley, P. D. (2012). Mitochondrial capture and
 1163 incomplete lineage sorting in the diversification of balitorine loaches (Cypriniformes,

1164 Balitoridae) revealed by mitochondrial and nuclear genes. *Zoologica Scripta*, 41(3),
1165 233–247.

1166 Tilak, M.-K., Justy, F., Debiais-Thibaud, M., Botero-Castro, F., Delsuc, F., & Douzery, E. J.
1167 P. (2015). A cost-effective straightforward protocol for shotgun Illumina libraries
1168 designed to assemble complete mitogenomes from non-model species. *Conservation*
1169 *Genetics Resources*, 7(1), 37–40.

1170 Van Dam, M. H., & Matzke, N. J. (2016). Evaluating the influence of connectivity and
1171 distance on biogeographical patterns in the south-western deserts of North America.
1172 *Journal of Biogeography*, 43(8), 1514–1532. <https://doi.org/10.1111/jbi.12727>

1173 Voris, H. K. (2000). Maps of Pleistocene sea levels in Southeast Asia: Shorelines, river
1174 systems and time durations. *Journal of Biogeography*, 27(5), 1153–1167.
1175 <https://doi.org/10.1046/j.1365-2699.2000.00489.x>

1176 Wallace, A. R. (1869). *The Malay Archipelago*. London: Harper.

1177 Wallis, G. P., Cameron□Christie, S. R., Kennedy, H. L., Palmer, G., Sanders, T. R., &
1178 Winter, D. J. (2017). Interspecific hybridization causes long□term phylogenetic
1179 discordance between nuclear and mitochondrial genomes in freshwater fishes.
1180 *Molecular Ecology*, 26(12), 3116–3127.

1181 Weir, J. T., & Schluter, D. (2007). The latitudinal gradient in recent speciation and extinction
1182 rates of birds and mammals. *Science*, 315, 1574–1576.

1183 Westerhold, T., Marwan, N., Drury, A. J., Liebrand, D., Agnini, C., Anagnostou, E., ...
1184 Florindo, F. (2020). An astronomically dated record of Earth's climate and its
1185 predictability over the last 66 million years. *Science*, 369(6509), 1383–1387.

1186 Wiens, J. J., & Donoghue, M. J. (2004). Historical biogeography, ecology and species
1187 richness. *Trends in Ecology and Evolution*, 19(12), 639–644.

1188 Woodruff, D. S. (2010). Biogeography and conservation in Southeast Asia: How 2.7 million
1189 years of repeated environmental fluctuations affect today's patterns and the future of the
1190 remaining refugial-phase biodiversity. *Biodiversity and Conservation*, 19(4), 919–941.

1191 <https://doi.org/10.1007/s10531-010-9783-3>

1192 Wurster, C. M., Rifai, H., Zhou, B., Haig, J., & Bird, M. I. (2019). Savanna in equatorial
1193 Borneo during the late Pleistocene. *Scientific Reports*, 9(1), 1–7.
1194 <https://doi.org/10.1038/s41598-019-42670-4>

1195 Zachos, J. C., Dickens, G. R., & Zeebe, R. E. (2008). An early Cenozoic perspective on
1196 greenhouse warming and carbon-cycle dynamics. *Nature*, 451(7176), 279–283.

1197 Zhang, J., Kapli, P., Pavlidis, P., & Stamatakis, A. (2013). A general species delimitation
1198 method with applications to phylogenetic placements. *Bioinformatics*, 29(22), 2869–
1199 2876. <https://doi.org/10.1093/bioinformatics/btt499>

1200

Biosketch

Arni Sholihah is a lecturer at Institut Teknologi Bandung (Indonesia), and obtained her PhD at the doctoral school GAIA (Université de Montpellier). She is interested in the biogeography and conservation of Southeast Asian biotas. The team consists in evolutionary biologists and/or specialists of Southeast Asian freshwater biotas interested in mechanisms of diversity build-up and conservation.

Editor: Jonathan Waters

Author contributions: AS, DW, LR and NH conceived the study. AS, TS, HD, DW, LR and NH conducted sampling. AS, EDT, JP, MKT, YF, JFA, LR and NH conducted the sequencing. AS, EDT, FC, LR and NH analysed the data. AS and NH drafted the initial version of the manuscript, and all authors further contributed to draft and approved the final version.

Appendices

Figure S1 Bayesian maximum clade credibility tree of Rasborinae based on 14 partitions (13 protein coding and one rRNA coding partitions), two prior clock rates (0.3% per Myrs for rRNA partition, 1.2% per Myrs for the 13 protein coding partitions) and a birth-death diversification model. Node circles indicate posterior probabilities (PP) and node bars represent 95% highest posterior density (HPD) of age estimates.

Figure S2 Lineage through time plots for each clade and distribution of the species and MOTUs age estimates (A), speciation rates through time according to BCST model for each clade (B), sea-levels (C, following Miller et al. 2005).

Figure S3 Geographic pattern of speciation in Rasborinae. The plots show the numbers of speciation events associated to the four geographic patterns of speciation as follows: between islands and within paleorivers, between islands and between paleorivers, within islands and within paleorivers, and within islands and between paleorivers. Each bar shows the number of speciation events for each category for all clades (A), Clade I (B), Clade II (C), Clade III (D), and Clade IV (E).

Figure S4 Mitochondrial MOTU trees reconstructed using StarBEAST2. A to D, Bayesian maximum credibility species trees for Clades I, II, III and IV, respectively, including posterior probabilities (PP, node circles) and 95% highest posterior density (HPD, node bars).

Table S1 List of DNA sequences used in present study, including systematic (Order, family, genus, species and updated species names), field numbers (BOLD), GenBank accession numbers for COI and Complete Mitochondrial Genomes (CMG), OTU numbers assigned, locality, geographic distribution (Locality, island and paleoriver) and phylogenetic status (ingroup, outgroup).

Table S2 Description of the 17 diversification models tested including acronym, rate variation, number of parameters and parameters involved (see material and methods section for details).

Table S3 Likelihood scores for each diversification model for each group. Description of each model can be found in Table S1.

Table S4 Log likelihood scores and Akaike Information Criterion values for DEC, DEC+J, DIV, DIV+J, BAY and BAY+J as calculated using BioGeoBEARS for island-based and paleoriver-based partitioning.

Figures and Tables captions

Figure 1 Geological reconstructions of the Indo-Australian archipelago since the middle Oligocene (modified from Lohman et al. 2011), paleoriver reconstruction in the Pleistocene (modified from Voris, 2000 and Woodruff, 2010) and sea levels fluctuations (adapted from Miller et al. 2005). A, middle Oligocene *ca.* 30 Ma. B, early Miocene *ca.* 20 Ma. C, late Miocene *ca.* 10 Ma. D, early Pliocene *ca.* 5 Ma. E, modern including limits of exposed land during -60 m and -120 m sea level drops, contour of the paleoriver watersheds and sampling sites. F, sea level reconstructions over the last 15 Myr.

Figure 2 Mechanisms of lineage dispersal and proliferation through eustasy (Paleoriver hypothesis). During interglacial times, sea-level highstands induce a fragmentation of watersheds into multiple tributaries further connected during sea-level lowstands of glacial times. *In situ* radiation panel depicts a scenario of watershed colonisation during glacial times

from an initial ancestor (a), further fragmented during sea-level highstand and resulting in a locally radiating lineage (a1-a8). Jump dispersal and *in situ* radiation panel depicts a scenario of watershed colonisation during glacial times from two ancestors (a, b), further fragmented during sea-level highstands and resulting in multiple local radiations (b1-b2, a1-a6).

Figure 3 Phylogenetic reconstructions in Rasborinae based on 79 mitogenomes with Rasborinae taxa in black, outgroups in grey and four major clades labelled Clade I to Clade IV. A, Bayesian maximum clade credibility tree based on 14 partitions (13 protein coding and one rRNA coding partitions), two prior clock rates (0.3% per Myrs for rRNA partition, 1.2% per Myrs for the 13 protein coding partitions) and a Yule diversification model. Node circles indicate posterior probabilities (PP) and node bars represent 95% highest posterior density (HPD) of age estimates. B, Maximum likelihood phylogenetic tree based on 14 partitions (13 protein coding and one rRNA coding partitions). Node circles indicate bootstrap proportion (BP).

Figure 4 Mitochondrial gene trees of Rasborinae Clades I, II, III and IV. Maximum likelihood trees and bootstrap proportions (BP, node circles) are in panels A, C, E, and G, or Clades I, II, III and IV, respectively. Bayesian maximum clade credibility trees, posterior probabilities (PP, node circles) and 95% highest posterior density (HPD, node bars) are in panels B, D, F, and H for Clades I, II, III, and IV, respectively. All ML trees were rooted using *A. mola*, *R. vaterifloris* and *H. atukorali*, Clade I ML tree (A) included additional extra-groups as follows: *R. cephalotaenia*, *R. einthovenii*, *R. sumatrana*, *R. dusonensis*, *R. aprotaenia*, *R. semilineata*, *R. rutteni*, *R. bankanensis*, and *R. borneensis*. Clade II ML tree (C) included additional extra-groups as follows: *R. cephalotaenia*, *R. einthovenii*, *R. sumatrana*, *R. dusonensis*, *R. aprotaenia*, *B. maculatus*, *T. n sp*, and *P. balaena*. Clade III ML tree (E) included additional extra-groups as follows: *B. maculatus*, *T. n sp*, *P. balaena*, *R. borneensis*, *R. bankanensis*, *R. semilineata*, *R. rutteni*, *R. sumatrana*, *R. dusonensis*, and *R. aprotaenia*. Clade IV ML tree (G) included additional extra-groups as follows: *B. maculatus*, *T. n sp*, *P. balaena*, *R. borneensis*, *R. bankanensis*, *R. semilineata*, *R. rutteni*, *R. cephalotaenia*, and *R. einthovenii*.

Figure 5 Mitochondrial ancestral area estimations of Rasborinae Clades I, II, III, and IV. A to D, ancestral area estimations for Clades I, II, III, and IV, respectively, according to an

1301 island-based (left) or paleoriver-based (right) geographical partitioning. Nodes labelling is
1302 temporally ordered and referred in the text.

1303
1304 **Figure 6** Dispersal pathways and *in situ* diversification among Sundaland Rasborinae
1305 between 31-17.5 Ma (A, Early Miocene), 17.5-7.5 Ma (B, Late Miocene) and 7.5 Ma-present
1306 (C, Early Pliocene). Arrows represent dispersal pathways according to ancestral area
1307 estimations (Table 2) and stars indicates *in situ* diversification.

1308
1309 **Table 1** Summary statistics of the most likely diversification models for Rasborinae Clades I,
1310 II, III, and IV including acronym, rate variation, number of parameters (NP), speciation rates
1311 (λ), corrected Akaike Information Criterion (AICc) and Akaike weight (AIC ω).

1312
1313 **Table 2** Freshwater dispersal pathways and associated clades through time.

1314
1315 **Table 3** Summary statistics of geographical patterns of lineage divergence for Rasborinae
1316 Clades I, II, III, and IV.

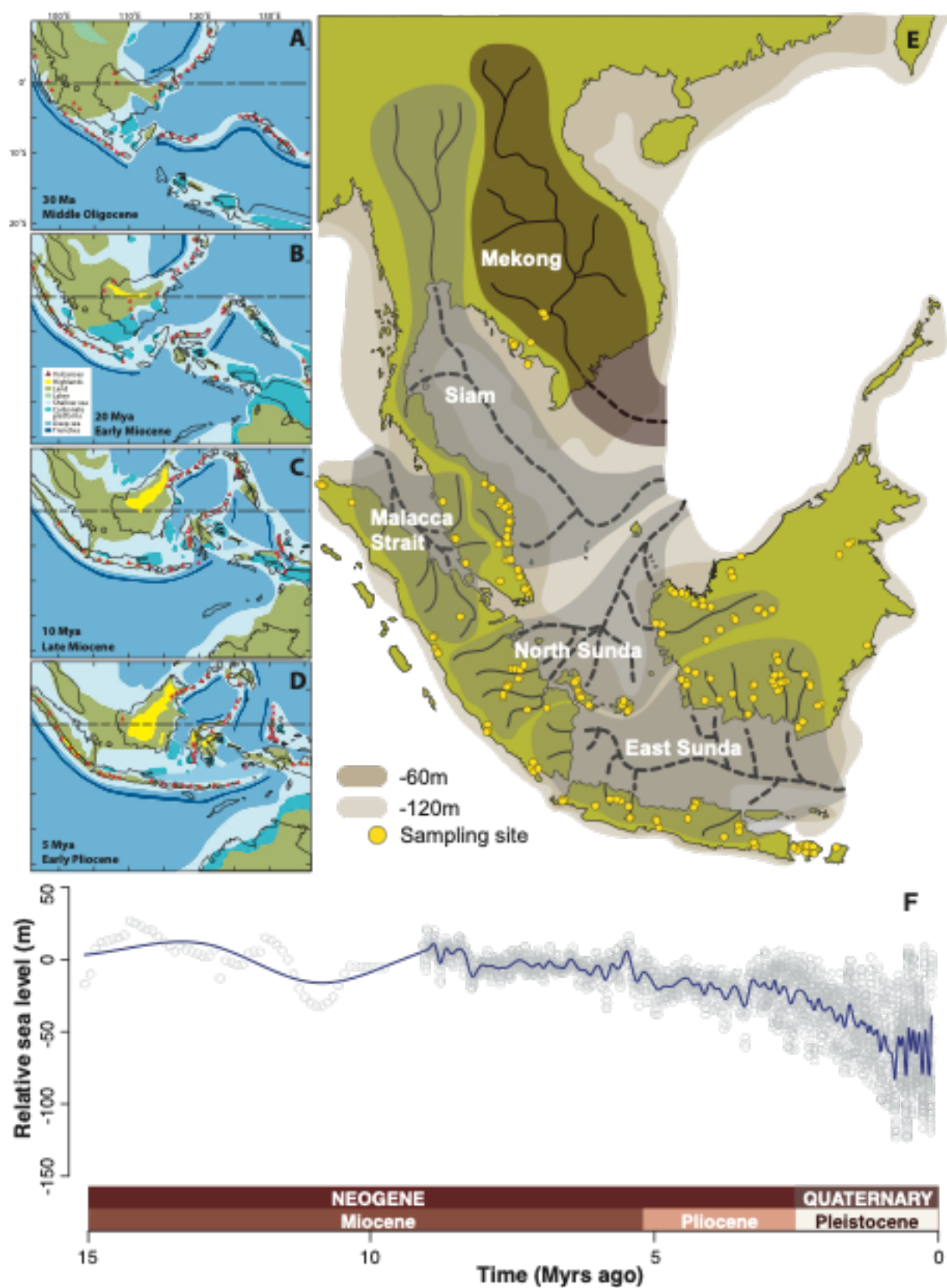


Fig. 1

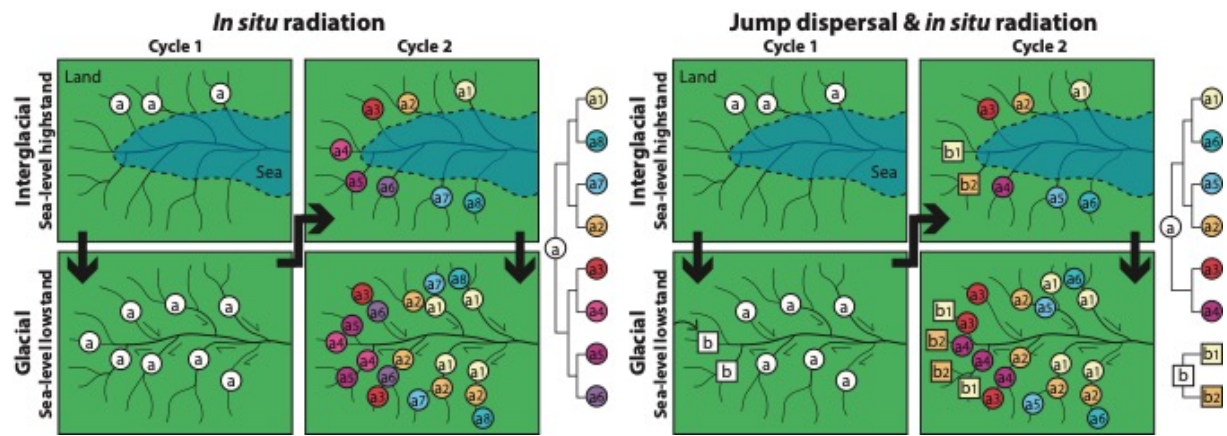


Fig. 2

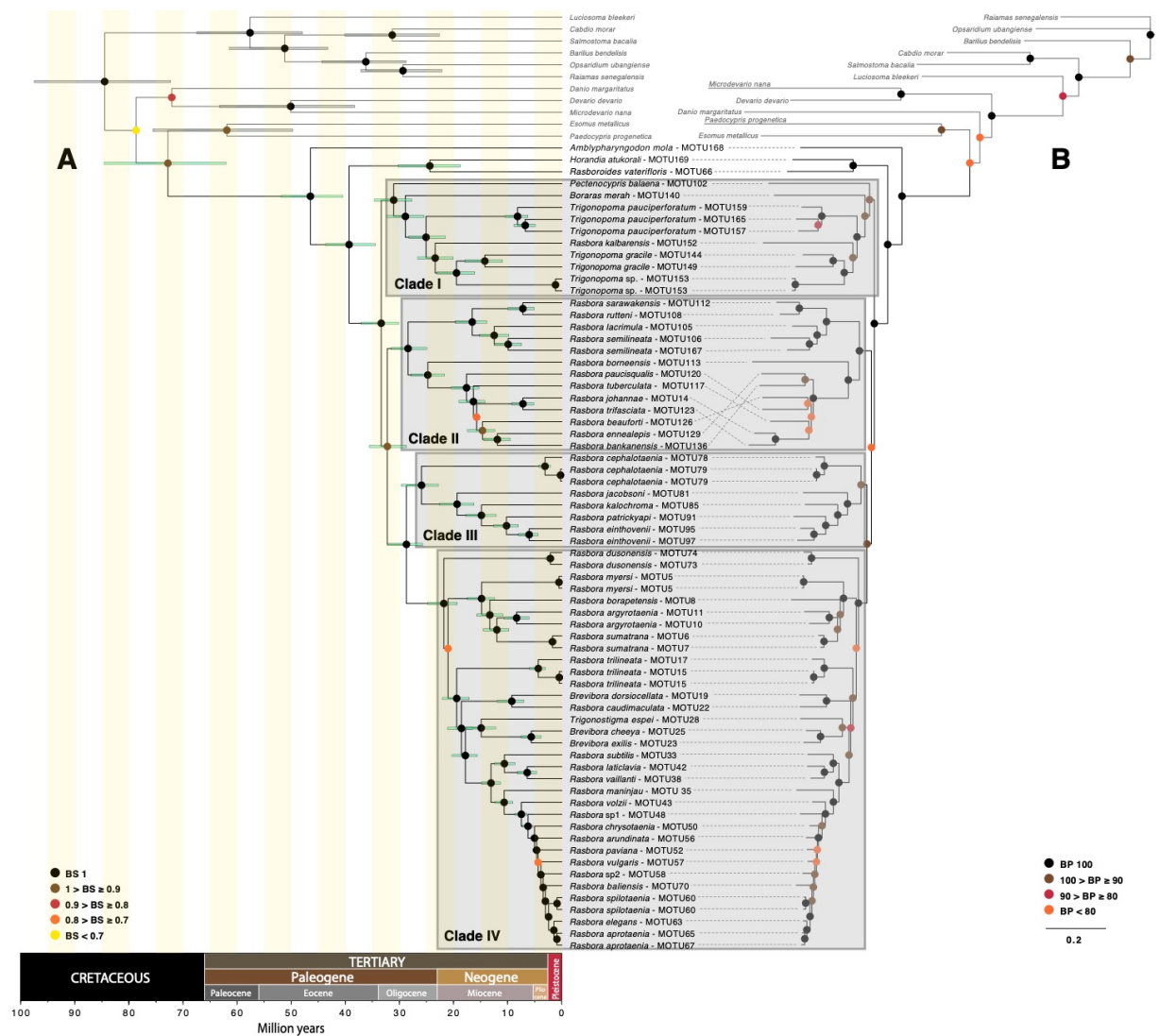


Fig. 3

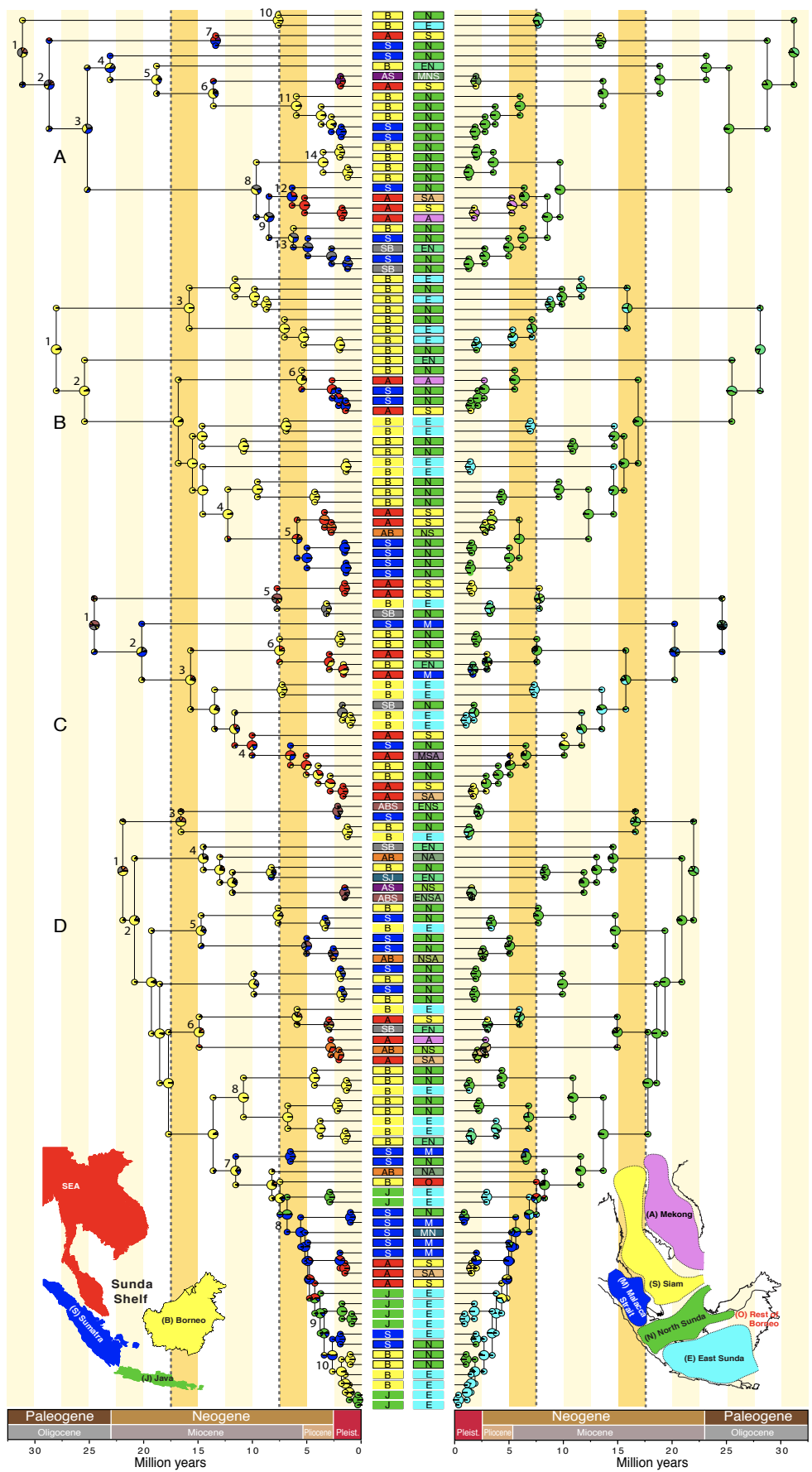


Fig. 5

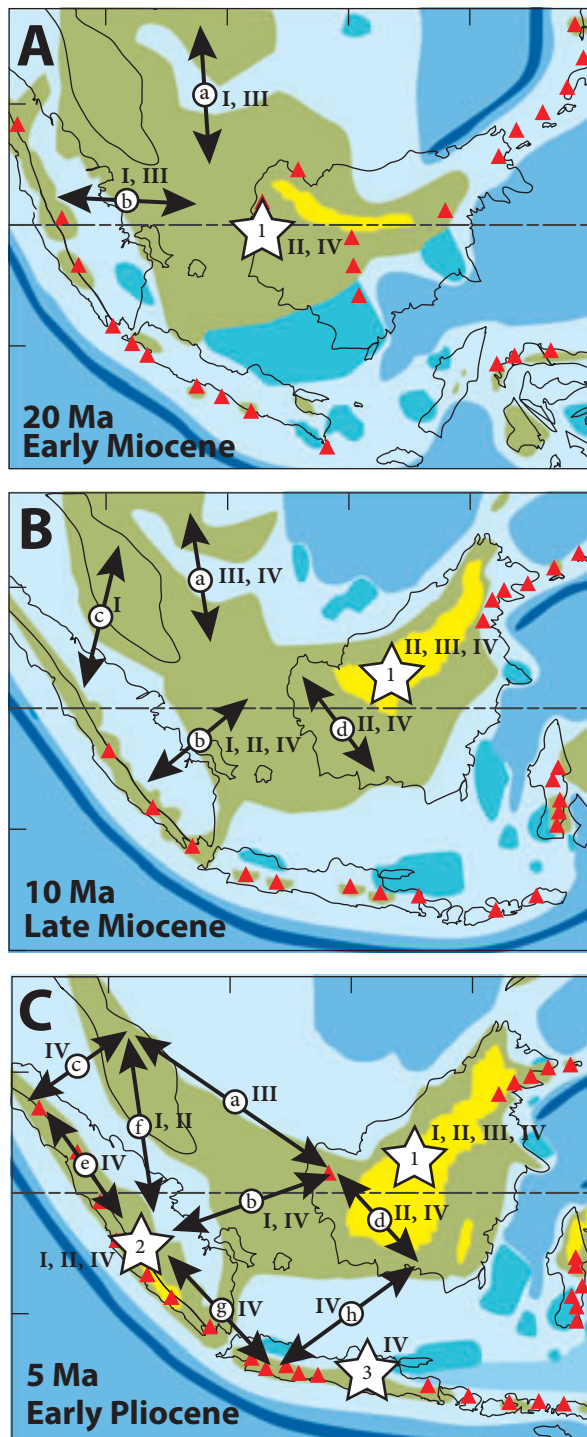


Fig. 6