# Paleo-Drainage Basin Connectivity Predicts Evolutionary Relationships across Three Southeast Asian Biodiversity Hotspots

Mark de Bruyn[1,*], Lukas Rüber[2], Stephan Nylinder[3], Björn Stelbrink[4], Nathan R. Lovejoy[5], Sébastien Lavoué[6], Heok Hui Tan[7], Estu Nugroho[8], Daisy Wowor[9], Peter K. L. Ng[7], M. N. Siti Azizah[10], Thomas von Rintelen[4], Robert Hall[11], and Gary R. Carvalho[1]

[1]*School of Biological Sciences, Environment Centre Wales, Bangor University, Deiniol Rd, Bangor LL57 2UW, UK;* [2]*Department of Vertebrates, Natural History Museum, Bernastrasse 15, CH-3005 Bern, Switzerland;* [3]*Department of Phanerogamic Botany, Natural History Museum, Svante Arrhenius vag 7, SE-114 18 Stockholm, Sweden;* [4]*Museum für Naturkunde, Leibniz-Institut für Evolutions-und Biodiversitätsforschung an der Humboldt-Universität zu Berlin, Invalidenstrasse 43, Berlin 10115, Germany;* [5]*Department of Biological Sciences, University of Toronto, 1265 Military Trail, Scarborough, ON M1C 1A4 Canada;* [6]*Department of Zoology, The Natural History Museum, Cromwell Rd, London SW7 5BD, UK;* [7]*Raffles Museum of Biodiversity Research, 21 Lower Kent Ridge Rd., National University of Singapore, Singapore 119077, Republic of Singapore;* [8]*Indonesian Research Institute for Freshwater Aquaculture, Jl. Sempur, Bogor, Java 16154, Indonesia;* [9]*Research Center for Biology (Puslit Biologi), Indonesian Institute of Sciences (LIPI), Jl. Raya Jakarta Bogor Km. 46, Cibinong 16911, Indonesia;* [10]*School of Biological Sciences, Universiti Sains Malaysia, Penang, Malaysia; and* [11]*Southeast Asia Research Group, Royal Holloway University, London, Egham, Surrey TW20 0EX, UK*
*Correspondence to be sent to: School of Biological Sciences, Environment Centre Wales, Bangor University, Bangor LL57 2UW, UK; E-mail: markus.debruyn@gmail.com.*

*Abstract*.—Understanding factors driving diversity across biodiversity hotspots is critical for formulating conservation priorities in the face of ongoing and escalating environmental deterioration. While biodiversity hotspots encompass a small fraction of Earth's land surface, more than half the world's plants and two-thirds of terrestrial vertebrate species are endemic to these hotspots. Tropical Southeast (SE) Asia displays extraordinary species richness, encompassing four biodiversity hotspots, though disentangling multiple potential drivers of species richness is confounded by the region's dynamic geological and climatic history. Here, we use multilocus molecular genetic data from dense multispecies sampling of freshwater fishes across three biodiversity hotspots, to test the effect of Quaternary climate change and resulting drainage rearrangements on aquatic faunal diversification. While Cenozoic geological processes have clearly shaped evolutionary history in SE Asian halfbeak fishes, we show that paleo-drainage re-arrangements resulting from Quaternary climate change played a significant role in the spatiotemporal evolution of lowland aquatic taxa, and provide priorities for conservation efforts. [Freshwater; geology; halfbeak; island radiation; Miocene; Pleistocene; river; Southeast Asia.]

Biodiversity hotspots have been mapped globally, encompassing species-rich regions of high conservation priority (Mittermeier et al. 2005). Although such regions encompass just 2.3% of Earth's surface, over 50% of the world's plant species and an estimated 42% of terrestrial vertebrate species are endemic to 34 hotspots worldwide (Mittermeier et al. 2005). Documenting species distributions and elucidating underlying drivers of speciation is a high priority as biodiversity hotspots and escalating threats often coincide (Joppa et al. 2011). Past climatic fluctuations, such as the glacial cycles of the Quaternary, are considered as primary determinants in the generation and distribution of biodiversity, both at the species and population levels (Hewitt 2000). Much of the temperate and boreal biomes were buried under ice during the Quaternary, with biota isolated in ice-free refugia, allowing associated faunal vicariant-isolation patterns to be investigated (Avise 2000; Hewitt 2000). While Quaternary climate has evidently shaped evolutionary relationships among biota from these regions, the role of the glaciations in driving diversification in the Tropics is less clear or has been dismissed entirely (Lessa et al. 2003; Esselstyn et al. 2009). Here, we show that while Cenozoic geological processes have clearly shaped evolutionary history in Southeast (SE) Asia, Quaternary climatic change played a significant role in the spatiotemporal evolution of

aquatic taxa examined across three of the four SE Asian tropical biodiversity hotspots.

SE Asia exhibits extraordinary species richness, yet it is under-studied and exposed to burgeoning conservation threats (Sodhi and Brook 2006). Four biodiversity hotspots have been defined in this region: Indo-Burma, Philippines, Sundaland, and Wallacea (Mittermeier et al. 2005). SE Asia's biota is currently likely in a "refugial-phase," as land above sea level occupies only 50–75% of what it did for the majority (> 90%) of the Quaternary (Woodruff 2010). Moreover, an expanding human population is increasing pressure on available habitat (Sodhi and Brook 2006). Understanding pattern and process in the dynamics of evolutionary novelty across this biodiversity-rich region is critical for establishing effective conservation strategies (Sodhi and Brook 2006; Lohman et al. 2011). Previous research examining the relative influences of geology versus Quaternary climate change on the generation of biodiversity in SE Asian fauna have, in general, remained inconclusive (Esselstyn et al. 2009; reviewed in Lohman et al. 2011). Similarly, since the early writings of Molengraaff and Weber (1921), the likely role of paleo-drainage systems on assembly of the SE Asian freshwater biota has been noted (reviewed in Voris 2000); however, attempts to investigate this empirically with a well-designed study have been lacking (see discussion in de Bruyn et al. 2012). No formal

comparative multispecies (phylogenetic) approach has yet been attempted, incorporating widespread sampling across multiple paleo-drainages to rigorously test alternative *a priori* hypotheses of diversification (e.g., Bohlen et al. 2011; see de Bruyn et al. 2012 for details).

Plate tectonic dynamics and past climatic change that drove eustatic sea levels have long made elucidating the major drivers of biodiversity in SE Asia a considerable challenge (Lohman et al. 2011; Gower et al. 2012). The region has an extremely dynamic geological (Hall 2011; reviewed in Lohman et al. 2011) and climatic history (Voris 2000; reviewed in Lohman et al. 2011); for example, the shallow SE Asian Sunda Shelf was above sea level for much of the Quaternary glacial periods—a region known today as "Sundaland" (Voris 2000; Lohman et al. 2011). This extended "land-bridge" likely facilitated dispersal of components of the biota between the SE Asian mainland and the Sunda Shelf islands (reviewed in Lohman et al. 2011). Temporal analytical approaches (e.g., based on ultrametric phylogenetic reconstructions) are thus required to segregate alternate processes— which have operated on different timescales—driving diversification (Ree and Smith 2008).

Significant recent advances have been made on the geological reconstruction of Cenozoic SE Asia (Hall 2011). Similarly, petroleum and other geological mapping enterprises across the now-drowned Sunda Shelf provided a framework for a series of maps illustrating the temporal distribution of Quaternary land and paleo-river drainage systems (Rainboth 1996; Voris 2000 and references therein). From these data, it is now possible to *a priori* formulate and test explicit expectations of evolutionary relationships among lowland freshwater taxa, based on both currently understood geological history (Hall 2011), and paleo-drainage re-arrangements resulting from Quaternary climatic change (Voris 2000).

We use multilocus molecular genetic data from dense multispecies sampling across three of the SE Asian hotspots, to test whether habitat re-arrangements arising from Quaternary climatic change drove diversification. We focus on three closely related genera (Lovejoy et al. 2004) of freshwater "halfbeak" fishes that are endemic to SE Asia: *Dermogenys*, *Nomorhamphus*, and *Hemirhamphodon*. All three SE Asian freshwater genera belong to the recently recognized family Zenarchopteridae (see discussion in Lovejoy et al. 2004), which also includes the SE Asian genus *Tondanichthys*, currently represented by only a single species (*Tondanichthys kottelati*, unavailable for inclusion), and the more widespread genus *Zenarchopterus*, which includes some saltwater tolerant taxa, included here as an outgroup genus (Lovejoy et al. 2004). Based on the predominantly marine life histories of the other four beloniform families (Lovejoy et al. 2004), ancestral marine dispersal in the distant past may have shaped the early evolutionary history of the family Zenarchopteridae. Six species of *Hemirhamphodon*, 12 species of *Dermogenys*, and 18 species of *Nomorhamphus* are currently recognized

(www.itis.gov), although relationships between genera *Dermogenys* and *Nomorhamphus* are somewhat unclear and in a state of flux (Lovejoy et al. 2004). While *Hemirhamphodon* and *Dermogenys* are codistributed and occur primarily (*Dermogenys*; also found on Sulawesi, i.e., Wallacea) or exclusively (*Hemirhamphodon*) to the west of the major biogeographical barrier known as Wallace's Line (i.e., across Indo-Burma, Sundaland, and the Philippines), *Nomorhamphus* is endemic to Sulawesi (and the Philippines, unavailable for inclusion) and provides a (presumably) long-isolated, closely related genus for comparison (Anderson and Collette 1991; Meisner 2001; Collette 2004). Specimens in *Hemirhamphodon* are larger (up to 120 mm) and more specialized, restricted primarily to forest streams, whereas in *Dermogenys* and *Nomorhamphus* they are smaller (up to 70 mm), with *Dermogenys* ubiquitous in many fresh- to brackish-water bodies, including the main bodies of the major rivers (Collette 2004; de Bruyn M., personal observation). All three genera are restricted to freshwaters (and brackish, in the case of *Dermogenys*) and are internally fertilized livebearers (viviparous), producing small numbers of relatively large young (Meisner 2001; Collette 2004). *Hemirhamphodon tengah*, in contrast, has been reported to be an egg layer (Dorn and Greven 2007). These life-history characteristics suggest limited dispersal capacity of larvae, compared, for example, with egg spawners, and thus biogeographical signal to be potentially high, that is, informative. Mapping of paleo-drainage distributions and the comparative phylogeographic approach allows for a powerful test of community assembly in response to Quaternary climatic change in one of the world's most biodiversity-rich tropical regions.

Paleo-drainage maps of SE Asia (e.g., Fig. 1) provide a testable framework for understanding the assembly of lowland freshwater faunas, and drives the sampling regime to incorporate both regions likely united via paleo-drainages in the past, and those that were not— that is essentially, negative controls (Rainboth 1996; Voris 2000). The availability of these "directly mapped" (geological) paleo-drainage maps provides a strong geographic basis for directly predicting connectivity, thereby reducing dependence on alternative models, such as ecological niche or other distributional models that are typically used to infer past global climate, and processes driving diversification, when past faunal habitats including temporal re-arrangements are unclear (e.g., Carnaval et al. 2009). If Quaternary paleo-drainage distributions, driven by climate-forced sea-level changes, determined the distribution of lowland freshwater diversity at the regional scale, phylogenetic diversity should be partitioned accordingly, reflecting Quaternary age. Similarly, inference on timing of paleo-drainage lineage-specific events (e.g., lineage splitting and growth) may help clarify our understanding of the evolution of SE Asian paleo-drainage systems (Rainboth 1996; Voris 2000). We use this exceptionally well-characterized paleo-landscape as a framework for examining the role of Quaternary climate change and
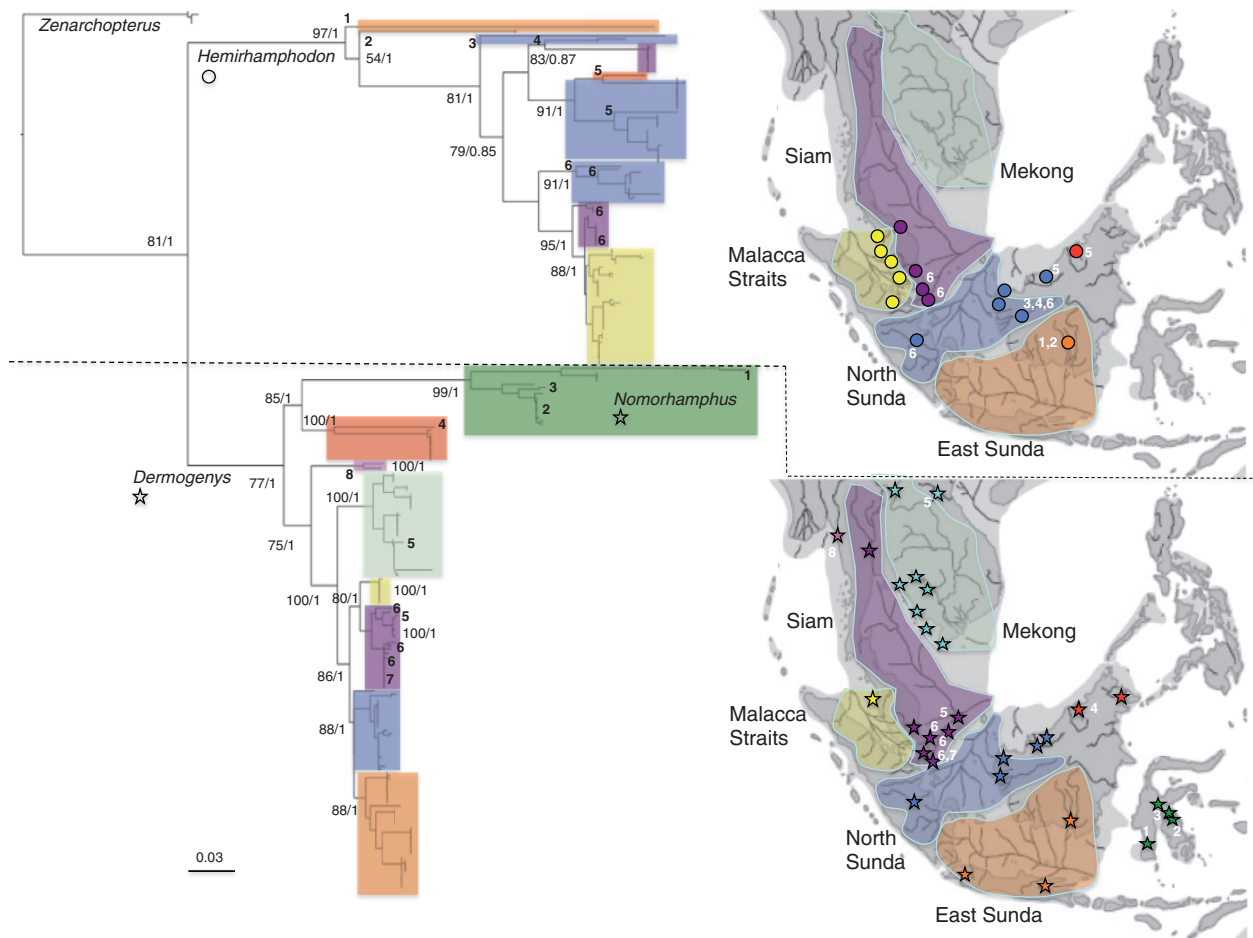
FIGURE 1.    Concatenated multigene phylogeny of SE Asian freshwater halfbeak fishes, with sampling locations superimposed on tree and over a paleo-drainage map of the region (Voris 2000) in respective shaded colors. Taxa are colored according to *a priori* paleo-drainage designation based on Voris (2000) (Supplementary Table S1). Maps reproduced by kind permission H. Voris and Field Museum, Chicago. Light gray shading shows land above sea level at maximal extent during the Quaternary, when sea levels were at −120 m compared with the present (Voris 2000). Present day geography illustrated in dark gray. *Zenarchopterus* was used to root the tree (Lovejoy et al. 2004). Top half of the phylogeny is *Hemirhamphodon* (shaded circles on map), while *Dermogenys* and *Nomorhamphus* (shaded stars) are shown below the dashed line. ML bootstrap values and Bayesian pps are indicated at nodes or alongside clades of interest. Individuals identified to species level by taxonomists are identified by bold numbers alongside branches (numbered 1–6 for six *Hemirhamphodon* taxa in upper half: 1=*H. chrysopunctatus*, 2=*H. tengah*, 3=*H. phaiosoma*, 4=*H. kapuasensis*, 5=*H. kuekenthali*, 6=*H. pogonognathus*; and 1–8 for eight *Nomorhamphus+Dermogenys* taxa in lower half: 1=*Nomorhamphus brembachi*, 2=*N. weberi*, 3=*N. megarrhamphus*, 4=*D. bruneiensis*, 5=*D. siamensis*, 6=*D. collettei*, 7=*D. pusilla*, 8=*D. burmanica*; with the corresponding number on the map to show sampling location). See Supplementary Table S1 for additional details. The same species identified in multiple reciprocally monophyletic lineages suggests either that taxonomy is not stable, for example, *H. pogonognathus*; *D. siamensis*; *D. collettei*; and *D. pusilla*, or misidentification by taxonomists.

associated drainage basin reconfigurations on assembly of the lowland freshwater fauna.

## MATERIALS AND METHODS

### Taxon Sampling

A total of 292 halfbeak individuals (excluding outgroups) were included in this study (Table S1; doi: 10.5061/dryad.23g44). The majority of fish were collected from the wild and preserved in ethanol, whereas some were sourced from collections. Tissue samples from fish identified to species level by taxonomists were also included in the analyses to examine taxonomic stability (Fig. 1; Supplementary Table S1). Voucher and tissue samples have been retained by the authors. To orient the ingroup phylogeny we included representatives of *Zenarchopterus* (*Zenarchopterus rasori*) as an outgroup (Lovejoy et al. 2004). Due to lack of available halfbeak samples, no specimens from the Philippines were included in this study.

### DNA—PCR and Sequencing

DNA was extracted using a standard high-salt DNA extraction protocol (Aljanabi and Martinez 1997). We

developed variable single-copy anonymous nuclear markers for the genera *Dermogenys* (de Bruyn et al. 2010a) and *Hemirhamphodon* (de Bruyn et al. 2010b), and trialled them on *Nomorhamphus*. Five loci (Dp5, Dp14, Dp21, Dp35, and Dp37) developed for *Dermogenys* amplified reliably across *Dermogenys* and *Nomorhamphus* samples, whereas three loci (Hp5, Hp54, and Hp56) developed for *Hemirhamphodon* amplified reliably across samples from this genus. We aimed to maximize the number of independent loci (Felsenstein 2006), rather than maximize the number of individuals genotyped across loci, thus, nuclear markers were amplified across a subset of individuals ($n=3$ per sampling site, plus all samples identified to species level by taxonomists), while mitochondrial control (CR) and *cytochrome oxidase I* (COI) regions were PCR amplified for all samples (Supplementary Tables S1 and S2). PCR conditions for nuclear markers are described in the primer notes (de Bruyn et al. 2010a, b). For mitochondrial markers, PCR was conducted as follows; each 25 µL PCR volume contained: 2 µL DNA extract, 1 U *Taq*, 1× buffer, 1.5 mM MgCl$^2$ (Invitrogen Ltd.; www.invitrogen.com), 0.2µM each primer, and 200µM each dNTP. CR primers used were: H16498 (5′-CCTGAAGTAGGAACCAGATG-3′) and L15995 (5′-CTCCACTATCAGCACCCAAAG-3′) (Meyer et al. 1990). COI primers used were: Fish-F2 (5′-TCGACTAATCATAAAGATATCGGCAC-3′) and Fish-R2 (5′-ACTTCAGGGTGACCGAAGAATCAGAA-3′) (Ward et al. 2005). PCR thermal cycles were set at: 95°C for 2 min, 35 cycles of 94°C for 45 s, 55°C for 45 s, and 72°C for 45 s, followed by a final extension at 68°C for 10 min. Multiple negative controls (no DNA) were used for all PCR runs. PCR products were purified using a Qiagen purification kit according to manufacturer's instructions, and subsequently sequenced at Macrogen (www.macrogen.com).

## Phylogenetic Analysis

*Alignment and model testing.*—Sequences from each gene were aligned using MUSCLE (Edgar 2004) on default settings and resulting matrices were subject to minor manual adjustments to remove apparent cases of false homologies. Model selection for best fitting substitution model for each alignment was conducted in MEGA v5.05 (Tamura et al. 2011) under the Bayesian information criterion (Schwarz 1978), which was preferred over the Akaike information criterion, to guard against over parameterization by averaging the likelihood over all included parameters. To test whether included genes evolve under the assumption of a strict molecular clock, each locus was set up in BEAST 1.7.2 (Drummond et al. 2012) under a basic coalescent model (Kingman 1982) while assigning each partition either a strict or a relaxed lognormal clock with vague but not uninformative priors on clock rates (exponential distributions with mean 0.1), and a fixed root of arbitrary height (100). Choice of clock models was made by Bayes factor (BF) calculations based on average values of the log marginal

likelihoods from dual runs for each partition obtained by stepping-stone (SS) sampling (Xie et al. 2011) (Table 1). Decisive support threshold for BF values was based on twice the natural logarithm (2 ln (BF) > 10) as suggested by Kass and Raftery (1995) of the null hypothesis (strict clock) over the alternative (relaxed clock). Each individual gene was tested for recombination using DualBrothers v1.1 (Minin et al. 2005) on default settings before further analysis to identify genes affected by processes not capturable by the coalescent models. All analyses performed in BEAST for this study were first run with empty alignments to check for cross-prior influence.

## Gene Trees, Species Tree Inference, and Time Calibrations

Individual phylogenies for each gene were estimated using (1) RAxML BlackBox (Stamatakis 2006; Stamatakis et al. 2008) maximum-likelihood (ML) bootstrapping under the GTR+Γ model (500 replicates) and (2) MrBayes v3.1.2 (Huelsenbeck and Ronquist 2001) conducting two independent runs of 10 million generations each, sampling every 10 000 trees, under the GTR+Γ (nst=6) substitution model.

Species tree analysis was performed using the multispecies coalescent module in BEAST 1.7.2 with specimens partitioned according to taxonomic rank (Heled and Drummond 2010). Genes were assigned a constant coalescent tree prior and best fitting clock and substitution models (Table 1), treating all but mitochondrial genes as unlinked groups. The species tree was assigned a pure birth (Yule) tree prior. Three geological calibrations, which putatively represent time to most recent common ancestors (TMRCA) of extant species or populations, were used. Using a normal distribution, we enforced constraints in the most conservative manner possible, interpreting the maximum recognized geological age of a geographical feature as the mean of a time horizon during which an internal bifurcation took place, with the 95% prior distribution representing soft bounds on the divergence time of that node. To further emphasize the uncertainty of the applied calibrations, none of the calibrated nodes had monophyly enforced. Geological calibrations were as follows:

1. TMRCA of the Surabaya *Dermogenys* population was assigned a normal distributed probability with mean 1.0 Ma and a standard deviation of 0.255 (prior credibility interval [PCI] of 0.5–1.5 Ma), which is consistent with the emergence of east Java (Hall 2011).

2. TMRCA of the representatives of the endemic Towuti species (*Nomorhamphus megarrhamphus*) (Collette 2004) was assigned a uniform estimate of 0.6 Ma (PCI 0.3–0.9 Ma), correlating to the probable age of Lake Towuti (Russell J., personal communication; Lohman et al. 2011).

TABLE 1. Choice of clock models made by BF calculations based on average values of the log marginal likelihoods from dual runs for each partition obtained by SS sampling (Xie et al. 2011)

| Gene | Clock model | Log marginal likelihood from SS estimate | Difference | BF | 2*ln(BF) | Log10BF | Preferred model if BF > 20 and/or log10 > 3 and/or 2*ln(BF) > 6 |
|---|---|---|---|---|---|---|---|
| HP5 | S | −790.995 | −2.026 | 7.589 | 4.053 | 0.880 | S |
| | LG | −793.023 | | 0.132 | | | |
| HP54 | S | −859.966 | −0.257 | 1.294 | 0.515 | 0.112 | S |
| | LG | −860.224 | | 0.772 | | | |
| HP56 | S | −784.237 | | 0.223 | | | S |
| | LG | −782.739 | −1.498 | 4.472 | 2.996 | 0.650 | |
| DP5 | S | −1580.933 | | 0.00000000617 | | | |
| | LG | −1562.029 | −18.903 | 162050432.170 | 37.806 | 8.209 | LG |
| DP14 | S | −825.737 | −20.013 | 491987790.428 | 40.027 | 8.692 | S |
| | LG | −845.751 | | 0.00000000203 | | | |
| DP21 | S | −1544.902 | | 0.0000193 | | | |
| | LG | −1534.050 | −10.852 | 51652.813 | 21.704 | 4.713 | LG |
| DP35 | S | −1704.422 | | 0.0000000351 | | | |
| | LG | −1687.256 | −17.165 | 28501703.455 | 34.331 | 7.454 | LG |
| DP37 | S | −1392.277 | | 0.000139 | | | |
| | LG | −1383.402 | −8.874 | 7150.225 | 17.749 | 3.854 | LG |
| COI DP | S | −4442.357 | | 0.00000104 | | | |
| | LG | −4428.587 | −13.769 | 955248.133 | 27.539 | 5.980 | LG |
| CR DP | S | −2681.818 | | 0.0517 | | | S |
| | LG | −2678.857 | −2.960 | 19.310 | 5.921 | 1.285 | |
| COI HP | S | −3985.201 | | 0.00000000000000136 | | | |
| | LG | −3950.971 | −34.223 | 734274824409785.009 | 68.459 | 14.865 | LG |
| CR HP | S | −2206.523 | | 0.0000534 | | | |
| | LG | −2196.685 | −9.837 | 18723.586 | 19.675 | 4.272 | LG |

3. TMRCA of *Nomorhamphus* populations restricted to Lake Matano was assigned the same age interval and probability as in calibration 2, although Lake Matano may be younger (Russell J., personal communication).

As these calibrations are optimally applicable to specific populations of certain species, they were assigned to the well-sampled mitochondrial data set rather than used as constraints on the estimated species tree. To test the relative degree of influence on speciation in *Dermogenys/Nomorhamphus* and *Hemirhamphodon*, two alternative species delimitations based on paleo-drainage lineages and present day taxonomy were set up as competing models for each group. Optimality was decided by SS sampling and comparing the average marginal likelihood score by BF calculations analogous to the one performed for modeltesting as described earlier. The explanatory power of the individually best scoring partition (i.e., the species delimitation with highest likelihood score, here used as the null hypothesis) for each of *Dermogenys/Nomorhamphus* and *Hemirhamphodon* was evaluated relative to their alternatives by BF comparisons. A threshold value of 2 ln (BF) > 10 was used as decisive support for the null against the alternatives (Table 2). All analyses in BEAST were run five times for 25 million generations each to

ensure independent convergence and chain mixing, and a relative proportion of trees was removed as burn-in from each run prior to summarizing. For all phylogenetic reconstructions, TRACER v1.5 (Drummond et al. 2012) was used to examine log files. Stationarity was assumed in BEAST runs when the effective sample size reached > 200 for all parameters (Drummond et al. 2006). Results from multiple independent runs were combined in LogCombiner v1.7.2 (Drummond et al. 2012), and maximum clade credibility (MCC) trees with median node heights were compiled in TreeAnnotator v1.7.2 (Drummond et al. 2012). All phylogenetic trees were visualized and edited in FigTree v1.3.1 (Rambaut 2006).

## Ancestral Range Reconstructions and Demographic Inference

Discrete Bayesian phylogeographic reconstructions were (Lemey et al. 2009) performed in BEAST on *Hemirhamphodon* and *Dermogenys/Nomorhamphus* mtDNA data separately, as combining the data sets for both *Dermogenys/Nomorhamphus* and *Hemirhamphodon* could result in suboptimal model performance due to the large number of geographic locations and inclusion of potentially deviating data points. To transfer the

TABLE 2.    Two alternative species delimitations based on paleo-drainage lineages and present day taxonomy set up as competing models for *Dermogenys/Nomorhamphus* and *Hemirhamphodon*

| Partition | | Log marginal likelihood from SS estimate | Difference | BF | 2 ln(BF) | Log10BF | Preferred model if BF > 20 and/or log10 > 3 and/or 2 ln(BF) > 6 |
|---|---|---|---|---|---|---|---|
| *DP paleo | Null | −12938.149 | | | | | *DP_paleo |
| *DP pres | | −13003.039 | −64.890 | Inf | Inf | Inf | |
| *HP paleo | Null | −8066.640 | | | | | *HP_paleo |
| *HP pres | | −8323.969 | −257.329 | Inf | Inf | Inf | |

Notes: Optimality was decided by SS sampling and comparing the average marginal likelihood score by BF calculations analogous to the one performed for model testing as described earlier. The explanatory power of the individually best scoring partition (i.e., the species delimitation with highest likelihood score, here used as the null hypothesis) for each of *Dermogenys/Nomorhamphus* and *Hemirhamphodon* was evaluated relative to their alternatives by BF comparisons. A threshold value of 2 ln (BF)> 10 was used as decisive support for the null against the alternatives.

effect of applied calibrations, we used derived posterior age distributions for the crown age of the outgroup from the *Dermogenys/Nomorhamphus* analysis as prior information for the *Hemirhamphodon* analysis, together with derived clock rates for the COI and CR genes. We thereby assume mutation rates for both data sets to be comparable and possible to model using secondary calibrations from a closely related sister group. Our decision to calibrate the *Hemirhamphodon* tree by application of secondary priors derived from the analysis of *Dermogenys/Nomorhamphus* was primarily motivated by technical reasons. Due to all geological priors being associated with shallow divergences in *Dermogenys/Nomorhamphus*, and the consistent lack of prior information on the tree root height, all node heights in *Hemirhamphodon* would be derived by traversing an unconstrained root. Priors placed at or near the root have the greatest potential to inform node height relative to all terminals, and should if available be preferred over terminally placed ones. Lack of calibration points at deeper nodes in the phylogeny allows the uncertainty associated with calibrating on short, shallow branch lengths to potentially propagate toward the root and decrease the explanatory power of the analysis. A combined approach of coestimating node heights in both *Dermogenys/Nomorhamphus* and *Hemirhamphodon* simultaneously using shallow node calibrations in *Dermogenys/Nomorhamphus* only, must in this context be considered sub-optimal and should best be avoided. For each genealogy, a simple constant rate coalescent tree model was preferred. Substitution models were linked across loci, and unlinked lognormal clock rate models were assigned to COI and CR data to allow each gene to individually contribute to branch lengths. For ancestral area reconstructions, each taxon was scored according to geographic sampling location ($K$) for present day locations and paleo-drainage lineages (e.g., Fig. 1). All pairwise distances between present day sampling locations were measured in kilometers, normalizing values by dividing the mean of all distances by each individual distance, and used as a prior for the relative probability of transitions between any pair of locations

in a symmetric rate matrix. In comparison, movement between paleo-drainage locations was assigned flat rate priors, indicating equal probability of transition between any two locations due to lack of comparable historical distance information. To maximize the effectiveness of the phylogeographic rate matrix, two measurements were taken (1) by assigning an exponential prior with mean 0.1 to the rate (clock) model for geographic traits, acting to minimize the number of transitions in the tree, and (2) minimizing significant rate transitions by use of the Bayesian Stochastic Search Variable Selection (BSSVS). Truncation of the BSSVS Poisson prior at $K$ may cause large rate matrices to suffer from increased rate multipliers, potentially elevating probabilities of less likely rates at the cost of more likely ones (Lemey et al. 2009), and would be observable by decreased state probabilities across all nodes in the tree, but due to a lack of alternative approaches we decided this was a reasonable strategy.

Each phylogeographic MCC tree was visualized in SPREAD (Bielejec et al. 2011) using decimal grade co-ordinates for each sampling location as input and generated as Google Earth files (.kml). As application of BSSVS in the rate matrices allow for only non-zero rates to be included, significantly supported rates can thus be identified using BF calculations. A lower boundary for significant rates was set to logBF > 3.0. Due to the effect of BSSVS on rate multipliers this can result in rates not present in the MCC trees to be identified as significant, and *vice versa*. However, an overall trend in which general areas are connected by increased number of significant rates should be interpreted as evidence of the historical connection between such areas (e.g., Supplementary Fig. S1).

To evaluate the use of a specific model for ancestral area inference we compared the phylogeographic inference in BEAST to ML methods as implemented in Lagrange (Ree et al. 2005; Ree and Smith 2008) and Mesquite v. 2.75 (Maddison and Maddison 2009). Lagrange setup was performed through the online configurator (http://www.reelab.net/lagrange/configurator/index)

as an unconstrained analysis with outgroup nodes excluded together with dispersal to/from outgroup ancestral state (Ree et al. 2005; Ree and Smith 2008), using the two individual MCC topologies from previous BEAST analyses of *Dermogenys/Nomorhamphus* and *Hemirhamphodon* data as input. ML estimation of ancestral areas in Mesquite was performed using the Markov k-state 1 parameter (Mk1) model (Lewis 2001) assigning the same probability to changes between any two states (equal to the JC model for substitution rates), and applied to the posterior distribution of trees derived from the Bayesian analysis of concatenated data in MrBayes as described earlier.

Demographic history was reconstructed for lineages that displayed strong evidence for a history dominated by paleo-drainage ("recent") connectivity (Fig. 1) across Sundaland, that is, *Dermogenys*, by integrating the model for extended Bayesian skyline plots (EBSPs) (Heled and Drummond 2008) in the Bayesian phylogeographic analyses as described earlier. Thus, inferred substitution rates are estimated directly in the analyses from the geological calibrations applied above. Ancestral population sizes were calculated for all lineages of *Dermogenys* corresponding to the postulated paleo-drainage areas of East Sunda, North Sunda, Mekong, Siam, and Malacca Straits (Fig. 1).

## RESULTS

### Sequence Data and Phylogenetic Relationships

Aligned matrices together comprise 2251 bp for the range of anonymous nuclear loci amplified for *Dermogenys/Nomorhamphus*. Individual alignments are of variable length (Dp5: 409 bp, Dp14: 324 bp, Dp21: 646 bp, Dp35: 487 bp, and Dp37: 385bp) containing variable proportions of informative sites (Dp5: 20.5%, Dp14: 6.5%, Dp21: 4.6%, Dp35: 17.2%, and Dp37: 15.1%). Corresponding lengths for anonymous nuclear loci for *Hemirhamphodon* are Hp5: 269 bp, Hp54: 262 bp, and Hp56: 351 bp with respective proportions of variable sites (Hp5: 8.3%, Hp54: 8.0%, and Hp56: 7.1%). Corresponding values for *Dermogenys/Nomorhamphus* mitochondrial genes are CR (345 bp, 33% informative) and COI (625 bp, 31% informative), and for *Hemirhamphodon* (375 bp, 31.2%) and (625 bp, 30.1%) respectively. All sequences used in this study are deposited in GenBank (accession numbers in Supplementary Table S2). Model testing for best fitting substitution model is presented in Supplementary Tables S3–S12. Prior BEAST analyses on empty alignments did not reveal any signs of cross-prior influence. All phylogenetic analyses recovered *Hemirhamphodon* and (*Dermogenys/Nomorhamphus*) as sister groups (Figs. 1 and 2; Supplementary Fig. S2), consistent with previous findings (Meisner 2001; Lovejoy et al. 2004). We did find support for *Nomorhamphus* as monophyletic (Figs. 1 and 2), although the genus was nested within a paraphyletic *Dermogenys* based on the concatenated data set (Fig. 1).

### Ancestral Range Reconstructions and Divergence Time Estimation

Bayesian and ML ancestral range analyses identified Borneo or Sulawesi as the likely region of origin of the *Dermogenys/Nomorhamphus* lineage (ultrametric mtDNA trees: Sulawesi = Bayesian state probability [B] 0.28, ML relative probability [ML] 0.28; Borneo = B 0.23, ML 0.16, all other locations an order of magnitude smaller; concatenated DNA tree: Sulawesi = ML 0.21; Borneo = ML 0.36, all other locations an order of magnitude smaller) (Supplementary Fig. S3 and Table S13). Analyses identified Borneo as the most likely origin of *Hemirhamphodon* (ultrametric mtDNA: B 0.30/0.28, ML 0.8745) (Supplementary Fig. S4 and Table S14). The timing of these events all coincides with the mid-Miocene 15–10 Ma (Figs. 2 and 3b,c). For further details on Bayesian (BEAST) and ML (Lagrange, Mesquite) ancestral range reconstruction and divergence time estimates, refer to Figure 2 and Supplementary Figures S1, S3, and S4, and Tables S13 and S14.

### Species Tree Analyses

BF evaluation of alternative species delimitations (Heled and Drummond 2010) unanimously favored paleo-drainage as the most influential speciation driver in both *Dermogenys/Nomorhamphus* and *Hemirhamphodon* (Table 2). Relationships between putative lineages in the paleo-drainage partitioned species tree of *Dermogenys/Nomorhamphus* describe a moderately supported topology grouping the lineages *Dermogenys bispina* and *D*.sp. "Brunei" as sister group to remaining taxa (Supplementary Fig. S5). The latter group, in turn, form a strongly supported clade with a posterior probability (pp) of > 0.95. Within this clade, a few lineages receive equally strong support. Most notably, the species containing specimens with an origin in the Siam (*D*. sp. "Siam") and Mekong (*D*.sp. "Mekong") paleo-basins form a sister pair (pp > 0.95), as do specimens originating from the North (*D*.sp. "North Sunda") and East Sunda (*D*.sp. "East Sunda") lineages, respectively. This clade, together with the paleo-lineages described as *Nomorhamphus megarrhampus* and *Nomorhamphus weberi*, together form another strongly supported clade, which in turn groups with the *D*. sp. "Siam"–*D*. sp. "Malacca"–*D*. sp. "Mekong" clade. The single yet unnamed species consisting of a number of specimens believed to belong to *Nomorhamphus* (*N*.sp.) show a strongly supported position as sister taxon to the latter clade. Corresponding support in *Hemirhamphodon* suggest a group of East Sunda paleo-drainage lineages to be the sister group to remaining species (Supplementary Fig. S6). Strongly supported clades include the sister relationship between
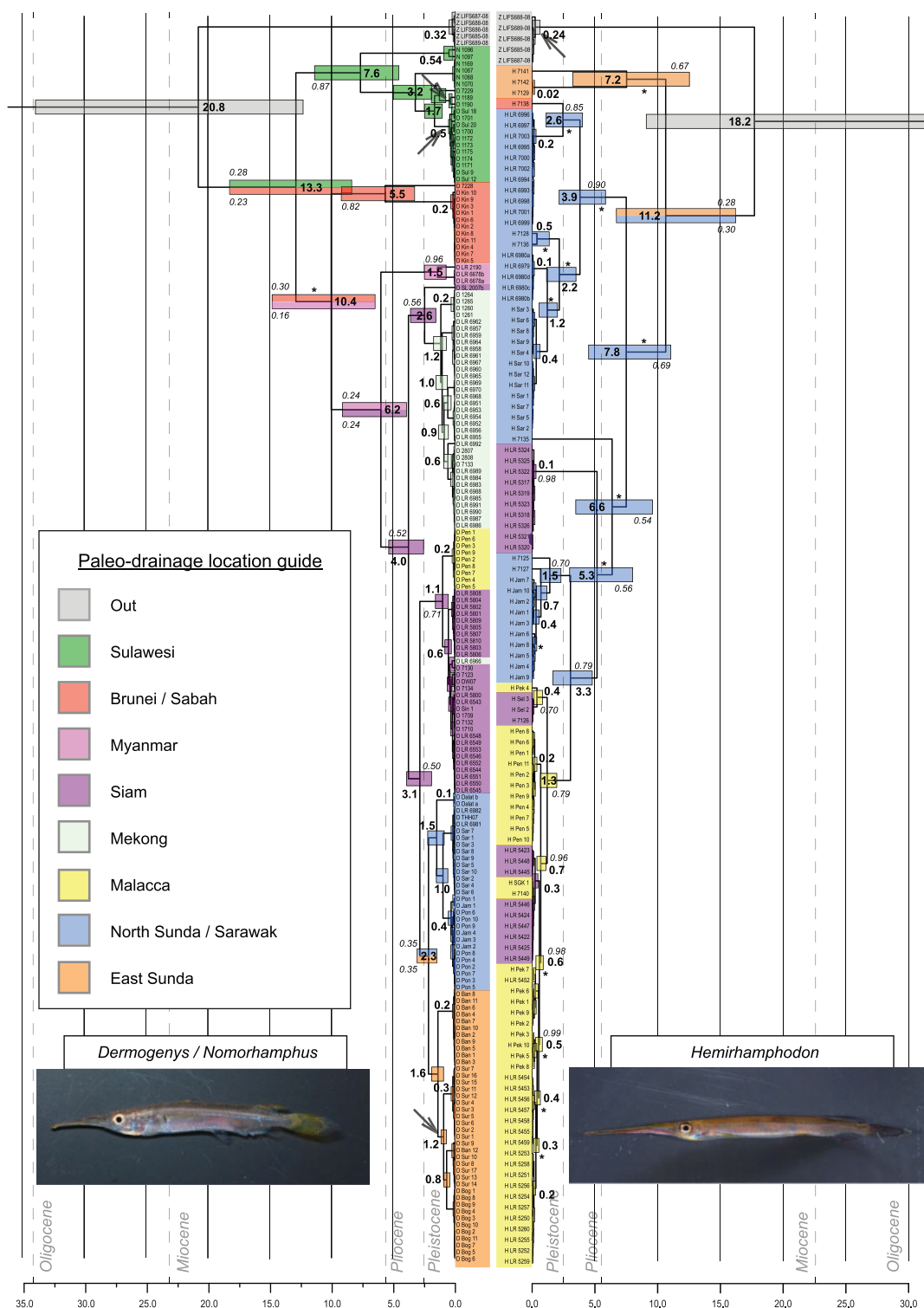
FIGURE 2.    Ultrametric Bayesian mitochondrial DNA trees of COI and control region variation in SE Asian freshwater halfbeak fishes. Taxa are colored according to *a priori* paleo-drainage designation based on Voris (2000) (Fig. 1; Supplementary Table S1). Values at nodes are median ages (in millions of years; bars=95% highest posterior densities). Shaded arrows at nodes of application indicate geological and secondary calibrations. Colors on bars indicate Bayesian node state (of origin), with relative state probability shown alongside. Pp at nodes <0.90 are indicated by asterisk. Photographs (by Heok Hui Tan) show examples of described species from the genera *Hemirhamphodon* (*H. pogonognathus*) and *Dermogenys* (*D. collettei*).
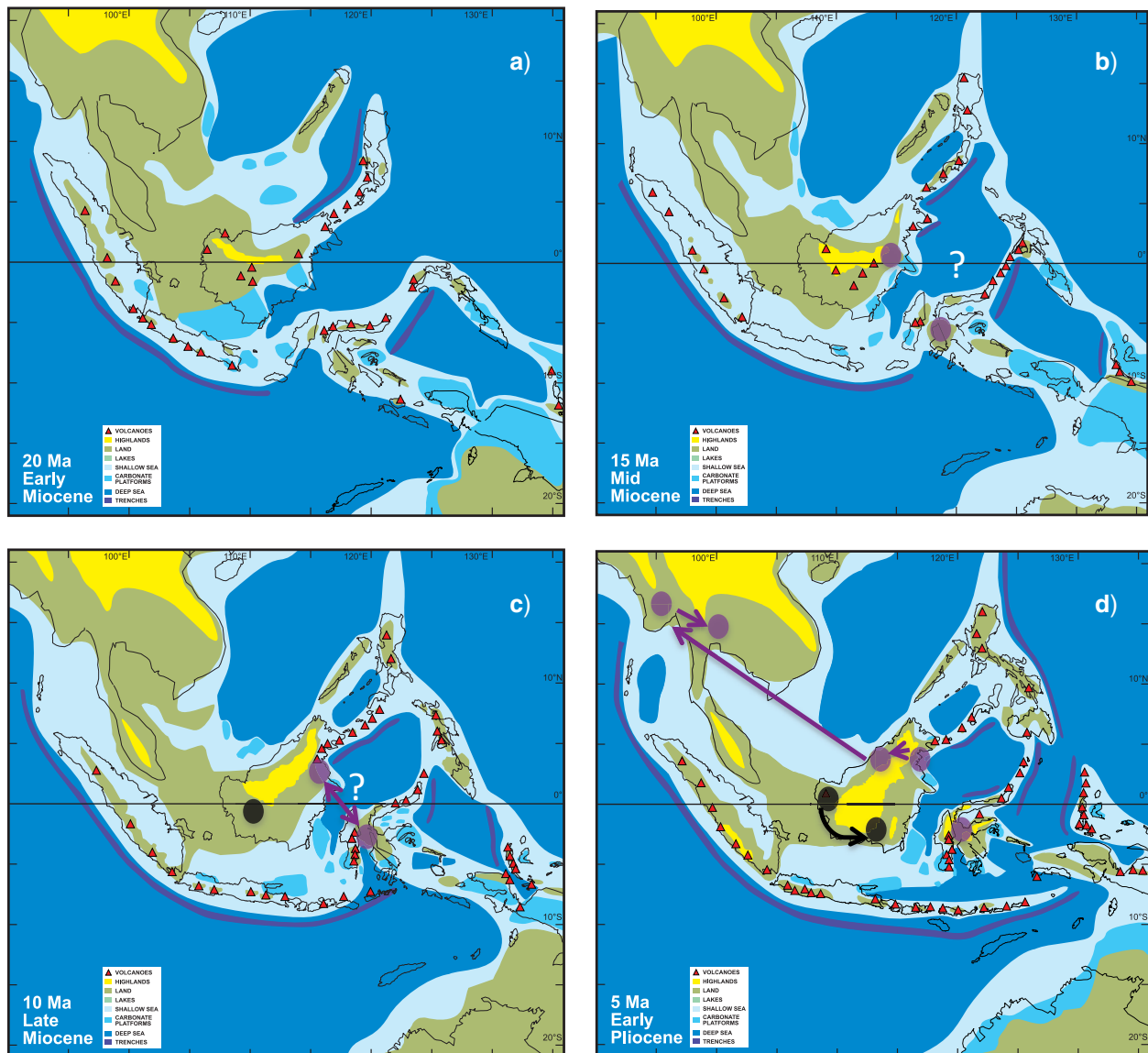
FIGURE 3.    Cenozoic model of early (pre-Quaternary) evolutionary history of SE Asian freshwater halfbeak fishes. Figure modified from Hall (2012). Purple circles = *Dermogenys + Nomorhamphus* (*Nomorhamphus* on Sulawesi only), Black = *Hemirhamphodon*. Reconstructions of paleogeography should be considered tentative working hypotheses at present, particularly for the Indochina region. a) SE Asia prior to inferred halfbeak evolution. b) (*Dermogenys + Nomorhamphus*) ancestral origin, either Borneo or Sulawesi. The Cenozoic radiation is consistent with a history of progressive range expansion across Sundaland, illustrated by colored arrows, followed by geological vicariance (mountain building/marine seaway opening) likely the major driver of "early" diversification (i.e., pre-Quaternary) (b–d). A sea-crossing (dispersal) from Borneo to Sulawesi c) or *vice versa* is inferred, however, as no land connection is likely based on Cenozoic geological reconstructions at this time (Hall 2011; Lohman et al. 2011).

"Kuching" and "North Sunda" specimens, plus the "Brunei" and "Mukah" specimens. Furthermore, these two clades together receive strong support as sister groups.

### Extended Bayesian Skyline Plots

Multigene EBSPs (Heled and Drummond 2008) recovered the demographic histories of some *Dermogenys* paleo-drainage lineages in considerable detail (Fig. 4). The East Sunda lineage (Fig. 4a) experienced an initial increase in effective population size between approximately 1.75–1.0 Ma, followed by a rapid and strong decline persisting to the present day. The Mekong and North Sunda lineages (Fig. 4b and c) seem to have retained stable population sizes up until the last 250 thousand years ago (Ka), when a decline occurred. These histories are contrary to the remaining two lineages (Siam and Malacca, Fig. 4d and e), where population sizes appear to have increased rapidly at two time intervals. For the Siam lineage, this increase began some 400 Ka and has shown steady growth until the present. The Malacca lineage remained stable up to
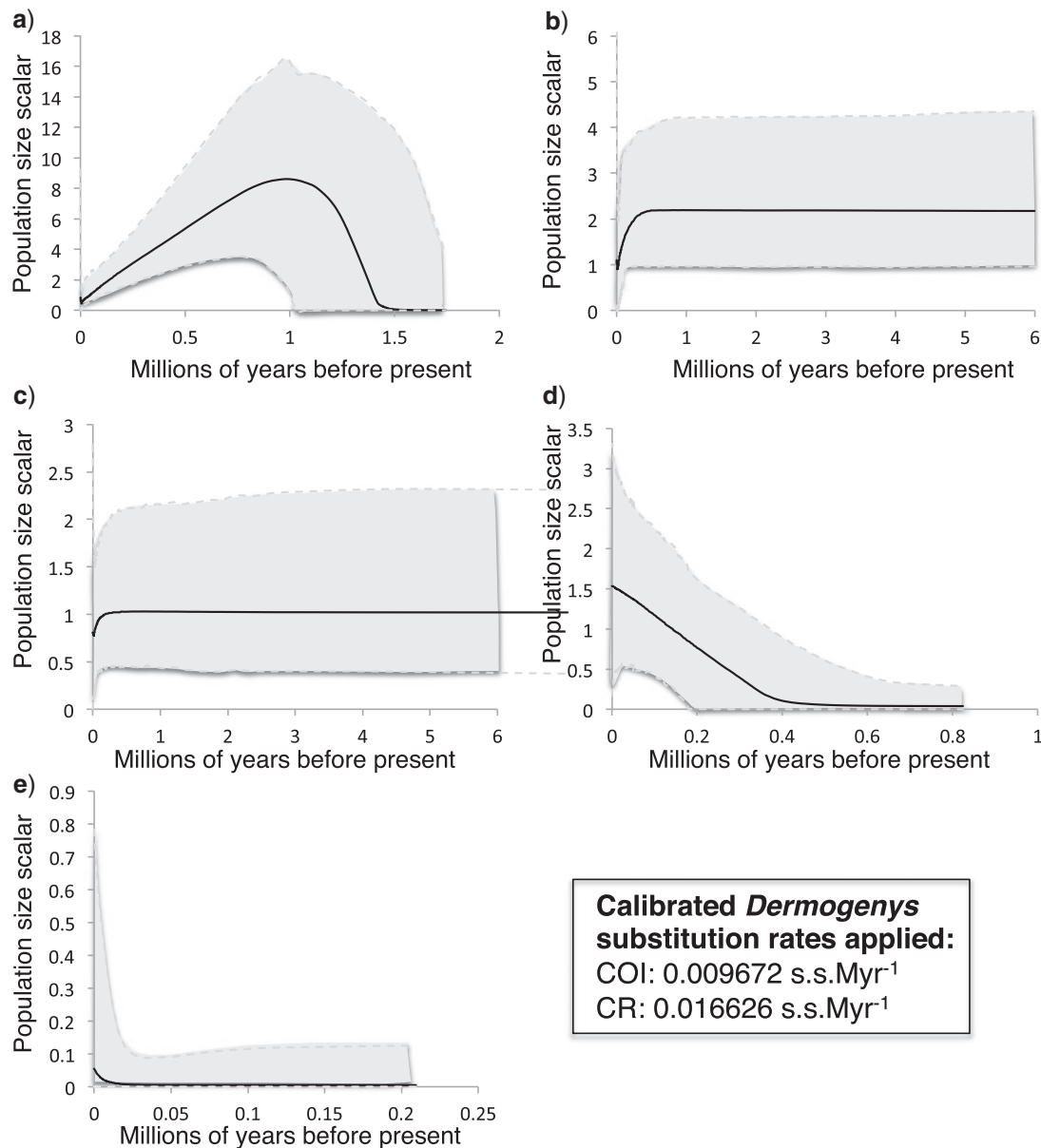
FIGURE 4. EBSPs (Heled and Drummond 2008) of the five *Dermogenys* paleo-drainage lineages, illustrating the chronology of demographic change, based on five nuclear and two mitochondrial genes. Population size scalar (in millions)/generation time = effective population size. Solid black line is the median effective population size, gray shaded area is the 95% highest posterior density intervals, incorporating phylogenetic and model uncertainty. Substitution rates are estimated directly in the analyses from the geological calibrations applied, and are consistent with published rates in other fishes (see Supplementary Table S1; Burridge et al. 2008). a) East Sunda, b) Mekong, c) North Sunda, d) Siam, and e) Malacca Straits. Note: Malacca Straits EBSP is based on only a single sampling site (Fig. 1).

approximately 25 Ka, and appears to have experienced an increasing effective population size since.

## DISCUSSION

The multigene phylogeny reflects both the chronology of Cenozoic geological evolution of SE Asia (Hall 2011), and Quaternary paleo-drainage (Voris 2000) connectivity well (Figs. 1–3; Supplementary Fig. S2), particularly for the more comprehensively sampled *Dermogenys*. Some 15 Ma, major geological changes were

taking place in SE Asia (Hall 2011; Lohman et al. 2011), with proto-Borneo and -Sulawesi emergent and increasing in size (Fig. 3). Based on the data we collected, our favored working hypothesis of the evolutionary history of these fishes follows. One caveat on this interpretation is the lack of availability of samples from the Philippines, and the unknown effect thereof on our analyses. Increased taxonomic coverage from currently unsampled regions is required in the future.

Ancestral *Dermogenys/Nomorhamphus* likely originated on Borneo or Sulawesi, whereas *Hemirhamphodon*

probably originated on Borneo according to our data (Fig. 3; Supplementary Figs. S1–S4 and Tables S13 and S14). *Dermogenys* subsequently crossed the Wallace Line from Borneo to Sulawesi or *vice versa* around 10 Ma (Fig. 3c; Supplementary Fig. S1). Interestingly, the split in the phylogeny leading to *Hemirhamphodon* and *Dermogenys* on Borneo ~10 Ma corresponds with the end of intense volcanic mountain building, which effectively split proto-Borneo north-south during the early Miocene (Figs. 1 and 3a–c; Hall 2011; Lohman et al. 2011). Minor range expansions leading to speciation events subsequently took place in both genera on proto-Borneo from ~10 to 5 Ma (Fig. 3d). Coincident with the start of the Pliocene, IndoBurma was colonized by *Dermogenys*, likely via the Sundaland land-bridge, effectively "seeding" the Siam paleo-drainage system, and what is today Myanmar (Fig. 3d; Supplementary Table S2). Subsequent extinctions or replacement of "intermediary" geographic lineages may have occurred. All other regions sampled for both genera are consistent with a Pleistocene radiation history, based on the three geological calibrations applied to the mitochondrial data (Fig. 2).

The concatenated phylogeny (Fig. 1), calibrated ultrametric mtDNA trees (Fig. 2), and BF analyses of phylogeographic diffusion rates (Lemey et al. 2009) (Supplementary Tables S15 and S16) indicated that halfbeaks sampled from adjacent drainage basins (e.g., Pekanbaru and Jambi, Sumatra), but representing distinct paleo-drainages (in this example, Malacca and North Sunda paleo-drainages, respectively [Voris 2000]), were phylogenetically dissimilar to one another, but related closely to halfbeaks from the same paleo-drainage on separate landmasses (in this example, Peninsula Malaysia and Borneo, respectively) (Fig. 1). Such relationships were consistent across all paleo-drainages for *Dermogenys* (41 of 46 significant BF of diffusion rates were between locations assigned *a priori* to the same paleo-drainage system; Supplementary Table S15; Figs. 1 and 2), but was less evident for *Hemirhamphodon* (1 of 14 significant BF rates within paleo-drainages; Supplementary Table S16), possibly reflecting either more limited sampling, or more likely, more specialized habitat requirements (i.e., forest-stream dependence [Collette 2004]) limiting paleo-dispersal.

In contrast to the geological drivers that effectively "seeded" halfbeaks on the major landmasses (Borneo/Sulawesi, IndoBurma) during the Cenozoic (Fig. 3; Supplementary Tables S13 and S14), the Bayesian and ML phylogeographic analyses (Fig. 2; Supplementary Fig. S1 and Tables S15 and S16) revealed that the Quaternary history of *Dermogenys*, at least, was consistent with a "radiation by paleo-drainage" model of diversification across SE Asia. Ancestral Cenozoic IndoBurma lineages: Myanmar, Siam—gave rise to derived paleo-drainage lineages: North-Sunda, East-Sunda, Malacca, and Mekong (Fig. 1). The chronology of this radiation advanced apparently in a stepwise southerly direction between adjacent paleo-drainages (Supplementary Fig. S1, only *Dermogenys* shown) (*Dermogenys*: Siam-Mekong BF=261.9, Siam-Malacca BF=16.3, North Sunda-East Sunda BF=7.9; *Hemirhamphodon*: North Sunda-Brunei BF=4.6, Siam-Malacca BF=35308.9), followed by radiation within the paleo-drainage system and subsequent diversification (cladogenesis), consistent with the retreat of marine conditions across Sundaland as sea-level dropped.

On a broader scale the comparison of timing of gene diversification versus speciation events in the ancestral-area optimized mitochondrial trees (Fig. 2) and the preferred (paleo-drainage, not current taxonomy) species trees for both *Dermogenys/Nomorhamphus* and *Hemirhamphodon* (Supplementary Figs. S5 and S6), indicate speciation to have occurred at much later dates than the geographic radiations. Such a result could indicate either extended times of conserved evolution of genes even after isolation of populations, presence of gene flow between ancient populations even long after establishment of geological barriers, or insufficient formation of dispersal barriers between populations. Species tree analyses also suggests that taxonomy of the study taxa may require revision (most notably, the validity of maintaining *Dermogenys/Nomorhamphus* as separate genera), and that paleo-drainage delineations, possibly preferable to current taxonomic delineations, may be a good starting point for such an analyses, particularly for *Dermogenys/Nomorhamphus* (Fig. 1; Supplementary Figs. S5 and S6).

The East Sunda lineage EBSP details both a significant Quaternary rapid growth phase, followed by an extreme prolonged bottleneck, that reduced the effective population size 10- to 15-fold by the end of the Pleistocene (Fig. 4a), most likely caused by sea-level rise and associated habitat reduction as Sundaland flooded (Voris 2000; Woodruff 2010), leading to the region's current geography. Whether each sea level rise and subsequent fall resulted in a loss of diversity, or whether this results from a single such event, is unclear. Molecular evidence supports not only the demographic histories of these lineages but also provides independent biological data to advance our understanding of the temporal evolution of SE Asia's paleo-drainage systems. While Sulawesi and Borneo and the IndoBurma/Sundaland paleo-drainage systems of Siam, North Sunda, and "Myanmar" contributed significantly to the radiation history of these fishes, Bayesian and ML analyses of ancestral states (Fig. 2; Supplementary Figs. S1 and S3 and Table S13) indicate that the Mekong, currently the region's major river system and producer of freshwater aquatic biomass, did not. Although this finding is somewhat contrary to expectation, geological data support this assertion, as the Mekong was likely a minor river system prior to the (possibly late) Pleistocene (Rainboth 1996). Our demographic inferences (EBSPs) suggest that the East Sunda paleo-drainage system was at its maximal extent some 1.5–1 Ma during the mid-Pleistocene (Fig. 4a), while the extended Siam paleo-drainage system has a

more recent Pleistocene history (< 400000 YBP), most likely too recent for detecting any molecular signal of effective population size decline at the level of molecular resolution utilized in this study (Fig. 4d). Interestingly, while molecular clock estimates are congruent with paleo-drainage connectivity during the Pleistocene (Fig. 2), this pre-dates the LGM, and appears to be driven largely by landscape change occurring during the early to mid-Pleistocene (e.g., Figs. 2 and 4). Questions arise as to why paleo-drainages connected halfbeaks during earlier Pleistocene glacial cycles, but appear not to have connected them during the last such cycle? Could incipient speciation have taken place prior to this? Further work is required to address these questions.

The lineage-specific EBSPs also demonstrate empirically that the majority of lineages examined are currently, as predicted for SE Asian taxa in general (Woodruff 2010), in a refugial phase with significantly smaller population sizes, and population connectivity, compared with the majority of the Quaternary (Fig. 4). Concomitantly, these halfbeak lineages are extremely vulnerable to localized stochastic extinction events (Kuussaari et al. 2009). Whether such decline and diminished population connectivity resulting from natural processes, alongside recent and escalating anthropogenic induced threats, have reduced population viability below the extinction threshold (Kuussaari et al. 2009) is unknown. At a broader level, we predict that similar significant losses of genetic diversity and population connectivity will apply across freshwater taxa from this region exposed to these refugial-phase dynamics (Woodruff 2010), making freshwater conservation efforts of immediate critical importance across SE Asia. Finally, the congruence between paleo-drainage basins and species and genetic diversity demonstrates that Pleistocene paleo-drainage systems are excellent initial candidates for guiding conservation efforts (*sensu* Avise's "Pleistocene Parks" concept [Avise 2008]), and for delineating freshwater (as opposed to terrestrial) biodiversity hotspots for which baseline data are critically required at present (Abell et al. 2008; Herbert et al. 2010), both within SE Asia and globally, where such paleo-drainage data exist.

## SUPPLEMENTARY MATERIAL

Data files and/or other supplementary information related to this paper have been deposited at Dryad data repository at http://datadryad.org, doi:10.5061/dryad.23g44, including Supplementary Tables S1–S16, Supplementary Figures S1–S7, 4x BEAST input files (.xml), 10x alignment files (.fas), while phylogenetic trees are available at TreeBASE (http://purl.org/phylo/treebase/phylows/study/TB2:S13840).

## FUNDING

## ACKNOWLEDGMENTS

## REFERENCES

Abell R., Thieme M.L., Revenga C., Bryer M., Kottelat M., Bogutskaya N., Coad B., Mandrak N., Contreras Balderas S., Bussing W., Stiassny M.L.J., Skelton P., Allen G.R., Unmack P., Naseka A., Ng R., Sindorf N., Robertson J., Armijo E., Higgins J.V., Heibel T.J., Wikramanayake E., Olson D., López H.L., Reis R.E., Lundberg J.G., Sabaj Pérez M.H., Petry P. 2008. Freshwater ecoregions of the world: a new map of biogeographic units for freshwater biodiversity conservation. BioScience 58:403–414.

Aljanabi S.M., Martinez I. 1997. Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. Nucl. Acids Res. 25:4692–4693.

Anderson W.D., Collette B.B. 1991. Revision of the freshwater viviparous halfbeaks of the genus *Hemirhamphodon* (Teleostei: Hemiramphidae). Ichthyol. Explor. Freshwat. 2:151–176.

Avise J.C. 2000. Phylogeography: the history and formation of species. Cambridge (MA): Harvard University Press.

Avise J.C. 2008. Three ambitious (and rather unorthodox) assignments for the field of biodiversity genetics. Proc. Natl. Acad. Sci. U.S.A. 105:11564–11570.

Bielejec F., Rambaut A., Suchard M.A., Lemey P. 2011. SPREAD: spatial phylogenetic reconstruction of evolutionary dynamics. Bioinformatics 20:2910–2912.

Bohlen J., Šlechtová V., Tan H.H., Britz R. 2011. Phylogeny of the Southeast Asian freshwater fish genus *Pangio* (Cypriniformes; Cobitidae). Mol. Phyl. Evol. 61:854–865.

Burridge C.P., Craw D., Fletcher D., Waters J.M. 2008. Geological dates and molecular rates: fish DNA sheds light on time dependency. Mol. Biol. Evol. 25:624–633.

Carnaval A.C., Hickerson M.J., Haddad C.F.B., Rodrigues M.T., Moritz C. 2009. Stability predicts genetic diversity in the Brazilian Atlantic forest hotspot. Science 323:785–789.

Collette B.B. 2004. Family Hemiramphidae Gill 1859. Annotated Checklists of Fishes. Halfbeaks, vol. 22. San Francisco (California): California Academy of Sciences. p. 1–35.

de Bruyn M., Grail W., Barlow A., Carvalho G.R. 2010a. Anonymous nuclear markers for SE Asian halfbeaks (genus *Dermogenys*). Conserv. Genet. Resour. 2:325–327.

de Bruyn M., Grail W., Carvalho G.R. 2010b. Anonymous nuclear markers for halfbeak fishes of the genus *Hemirhamphodon*. Conserv. Genet. Resour. 3:155–157.

de Bruyn M., von Rintelen T., von Rintelen K., Mather P.B., Carvalho G.R. 2012. Molecular biogeography and phylogeography of the freshwater fauna of the Indo-Australian Archipelago. In: Gower D.J., Johnson K., Richardson J., Rosen B., Rüber L., Williams S., editors. Biotic evolution and environmental change in Southeast Asia. Cambridge: Syst. Assoc. Vol., Cambridge University Press.

Dorn A., Greven H. 2007. Some observations on courtship and mating of *Hemirhamphodon tengah* Anderson & Collette, 1991 (Zenarchopteridae). Bull. Fish Biol. 9:99–104.

Drummond A.J., Ho S.Y.W., Phillips M.J., Rambaut A. 2006. Relaxed phylogenetics and dating with confidence. PLoS Biol. 4:e88.

Drummond A.J., Suchard M.A., Xie D., Rambaut A. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol. Biol. Evol. 29:1969–1973.

Edgar R.C. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5: 1–19.

Esselstyn J.A., Timm R.M., Brown R.M. 2009. Do geological or climatic processes drive speciation in dynamic archipelagos? The tempo and mode of diversification in Southeast Asian shrews. Evolution 63:2595–2610.

Felsenstein J. 2006. Accuracy of coalescent likelihood estimates: do we need more sites, more sequences, or more loci? Mol. Biol. Evol. 23:691–700.

Gower D.J., Johnson K., Richardson J., Rosen B., Rüber L., Williams S., editors. 2012. Biotic evolution and environmental change in Southeast Asia. Cambridge: Syst. Assoc. Vol., Cambridge University Press.

Hall R. 2011. Australia-SE Asia collision: plate tectonics and crustal flow. In: Hall R., Cottam M.A., Wilson M.E.J., editors. The SE Asian gateway: history and tectonics of Australia-Asia Collision. London: Geological Society of London Special Publication. p. 75–109.

Hall R. 2012. Sundaland and Wallacea: geology, plate tectonics and palaeogeography. In: Gower D.J., Richardson J.E., Rosen B.R., Rüber L., Williams S.T., editors. Biotic evolution and environmental change in Southeast Asia. Cambridge: Cambridge University Press. p. 32–78.

Heled J., Drummond A.J. 2008. Bayesian inference of population size history from multiple loci. BMC Evol. Biol. 8:289.

Heled J., Drummond A.J. 2010. Bayesian inference of species trees from multilocus data. Mol. Biol. Evol. 27:570–580.

Herbert M.E., McIntyre P.B., Doran P.J., Allan D.J., Abell R. 2010. Terrestrial reserve networks do not adequately represent aquatic ecosystems. Conserv. Biol. 24:1002–1011.

Hewitt G. 2000. The genetic legacy of the Quaternary ice ages. Nature 405:907–913.

Huelsenbeck J.P., Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17:754–755.

Joppa L.N., Roberts D.L., Myers N., Pimm S.L. 2011. Biodiversity hotspots house most undiscovered plant species. Proc. Natl. Acad. Sci. U.S.A. 108:13171–13176.

Kass R.E., Raftery A.E. 1995. Bayes factors. J. Am. Stat. Assoc. 430: 773–795.

Kingman J.F.C. 1982. The coalescent. Stoch. Process Appl. 13:235–248.

Kuussaari M., Bommarco R., Heikkinen R.K., Helm A., Krauss J., Lindborg R., Öckinger E., Pärtel M., Pino J., Rodà F., Stefanescu C., Teder T., Zobel M, Steffan-Dewenter I. 2009. Extinction debt: a challenge for biodiversity conservation. Trends Ecol. Evol. 24: 564–571.

Lemey P., Rambaut A., Drummond A.J., Suchard M.A. 2009. Bayesian phylogeography finds its roots. PLoS Comp. Biol. 5:e1000520.

Lessa E.P., Cook J.A., Patton J.L. 2003. Genetic footprints of demographic expansion in North America, but not Amazonia, during the Late Quaternary. Proc. Natl. Acad. Sci. U.S.A. 100: 10331–10334.

Lewis P.O. 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. Syst. Biol. 50:913–925.

Lohman D., de Bruyn M., Page T., von Rintelen K., Hall P.K.L., Shih H-T., Carvalho G.R., von Rintelen T. 2011. Biogeography of the Indo-Australian Archipelago. Ann. Rev. Ecol. Evol. Syst. 42:205–226.

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

Maddison W.P., Maddison D.R. 2009. Mesquite: a modular system for evolutionary analysis. Version 2.72 [Internet]. Available from: URL http://mesquiteproject.org.

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

Meyer A., Kocher T.D., Basasibwaki P., Wilson A.C. 1990. Monophyletic origin of Lake Victoria cichlid fishes suggested by mitochondrial DNA sequences. Nature 347:550–553.

Minin V.N., Dorman K.S., Fang F., Suchard M.A. 2005. Dual multiple change-point model leads to more accurate recombination detection. Bioinformatics 21:3034–3042.

Mittermeier R.A., Gil P.R., Hoffman M., Pilgrim J., Brooks T., Mittermeier C.G., Lamoreux J., da Fonseca G.A.B. 2005. Hotspots revisited: earth's biologically richest and most endangered terrestrial ecoregions. Washington: Conservation International.

Molengraaff G.A.F., Weber M. 1921. On the relation between the Pleistocene glacial period and the origin of the Sunda Sea (Java and South China-Sea), and its influence on the distribution of coral reefs and on the land and freshwater fauna. K. Ned. Akad. Wet. Verhan. 23:395–439.

Rainboth W. J. 1996. The taxonomy, systematics and zoogeography of *Hypsibarbus*, a new genus of large barbs (Pisces, Cyprinidae) from the rivers of southeastern Asia. Univ. Calif. Pubs. Zool. 129:xxiii–199.

Rambaut A. 2006. FigTree. v1.3.1 [Internet]. Available from: URL http://tree.bio.ed.ac.uk/software/figtree.

Ree R.H., Moore B.R., Webb C.O., Donoghue M.J. 2005. A likelihood framework for inferring the evolution of geographic range on phylogenetic trees. Evolution 59:2299–2311.

Ree R.H., Smith S.A. 2008. Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. Syst. Biol. 57:4–14.

Schwarz G.E. 1978. Estimating the dimension of a model. Ann. Stat. 6:461–464.

Sodhi N.S., Brook B.W. 2006. Southeast Asian biodiversity in crisis. Cambridge: Cambridge University Press.

Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22:2688–2690.

Stamatakis A., Hoover P., Rougemont J.A. 2008. A rapid bootstrap algorithm for the RAxML web servers. Syst. Biol. 57:758–771.

Tamura K., Peterson D., Peterson N., Stecher G., Nei M., Kumar S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28:2731–2739.

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

Ward R.D., Zemlak T.S., Innes B.H., Last P.R., Hebert P.D.N. 2005. DNA barcoding Australia's fish species. Philos. Trans. R. Soc. Lond. B. 360:1847–1857.

Woodruff D. 2010. Biogeography and conservation in Southeast Asia: how 2.7 million years of repeated environmental fluctuations affect today's patterns and the future of the remaining refugial-phase biodiversity. Biodivers. Conserv. 19:919–941.

Xie W., Lewis P.O., Fan Y., Kuo L., Chen M.-H. 2011. Improving marginal likelihood estimation for Bayesian phylogenetic model selection. Syst. Biol. 60:150–160.