

Synchronous diversification of parachuting frogs (Genus *Rhacophorus*) on Sumatra and Java

Kyle A. O'Connell^{a,b,*}, Amir Hamidy^c, Nia Kurniawan^d, Eric N. Smith^{a,b}, Matthew K. Fujita^{a,b}

^a Department of Biology, The University of Texas at Arlington, Arlington, TX 76019, USA

^b The Amphibian and Reptile Diversity Research Center, University of Texas at Arlington, Arlington, TX 76010, USA

^c Research and Development Center for Biology, Indonesian Institute of Science (LIPI), Widyasatwalo Building, Cibinong 16911, West Java, Indonesia

^d Department of Biology, Universitas Brawijaya, Jl. Veteran, Malang 65145, East Java, Indonesia

ARTICLE INFO

Keywords:

Allopatry
Divergence
Indonesia
Phylogeography
Sunda Shelf

ABSTRACT

Geological and climatological processes can drive the synchronous diversification of co-distributed species. The islands of Sumatra and Java have experienced complex geological and climatological histories, including extensive sea-level changes and the formation of valleys between northern, central, and southern components of the Barisan Mountain Range, which may have promoted diversification of their resident species. We investigate diversification on these islands using 13 species of the parachuting frog genus *Rhacophorus*. We use both mitochondrial and nuclear sequence data, along with genome-wide SNPs to estimate phylogenetic structure and divergence times, and to test for synchronous diversification. We find support for synchronous divergence among sister-species pairs from Sumatra and Java ~9 Ma, as well as of populations of four co-distributed taxa on Sumatra ~5.6 Ma. We found that sister species diverged in allopatry on Sumatra and conclude that divergence on Sumatra and Java was affected by sea-level fluctuations that promoted isolation in allopatry.

1. Introduction

Biotic responses to climatological or geological changes often drive diversification on tropical islands (Esselstyn et al., 2009). Climatic fluctuations can accelerate diversification by isolating species into refugia or by expanding suitable habitat, thus promoting dispersal (Nater et al., 2015). Likewise, geological changes can initiate diversification by isolating populations in allopatry. The Sunda Shelf (Sumatra, Java, Borneo, and the Malay Peninsula) has experienced a turbulent geological and climatological history from the Miocene to present (Lohman et al., 2011). Sumatra in particular has experienced dynamic tectonic processes, volcanism, dramatic surrounding sea-level changes, and extensive connectivity with surrounding landmasses during the Pleistocene (Hall, 2001, 2002, 2009, 2011, 2012a, 2012b, Lohman et al., 2011). For most of the past 25 million years (Ma), highland habitats on Sumatra have remained tropical, while lowland forests were frequently inundated by marine incursions, and also experienced extensive cooling and drying (Hall, 2009, 2012a).

Although past studies have largely focused on the role of Pleistocene sea-level fluctuations on diversification on Sumatra and Java, few studies have investigated the role of Miocene-Pliocene sea-level changes, or of the formation of physical barriers during this time period (Inger &

Voris, 2001; Leonard et al., 2015; Voris, 2000). During much of the Miocene, Sumatra was composed of several islands, with marine incursions serving as barriers to dispersal (van Bemmelen, 1949; Meijaard, 2004; Hall, 2012a). From the early Miocene to ~15 Ma, raised shorelines persisted on Sumatra, transforming volcanic peaks into small islands (Baumann, 1982; Haq et al., 1987; Batchelor, 1979; Anderson et al., 1993; Collins et al., 1995; Lourens & Hilgen, 1997; Barber et al., 2005). From 14 to 9 Ma sea levels receded, permitting dispersal between previously isolated volcanic islands (Batchelor, 1979; Baumann, 1982; Haq et al., 1987; Morley, 1998). This cycle continued, with sea levels rising from 8.5 to 6 Ma, receding from 5.8 to 5.4 My, and again rising from 5 to 4 My (Baumann, 1982; Haq et al., 1987; Krantz, 1991; Anderson et al., 1993; Van den Bergh et al., 2001).

Furthermore, van Bemmelen (1949) hypothesized the persistence of two transverse inland seaways on Sumatra from the early Miocene onward that divided Sumatra between the northern and central components of the Barisan Mountain Range (in the Padang Sidempuan Valley, just south of the Asahan High, which was an elevated region that ran transverse to the Sumatran mainland), and between the Gumai and Garba Mts (in the Pagar Alam Valley, Fig. 1). These seaways formed in the early Miocene, and completely subsided only in the middle Pliocene due to Barisan Mountain uplift (van Bemmelen, 1949). As

* Corresponding author at: Department of Vertebrate Zoology, National Museum of Natural History, Smithsonian Institution, Washington, DC, 20560, USA.
E-mail address: kyle.oconnell@mavs.uta.edu (K.A. O'Connell).

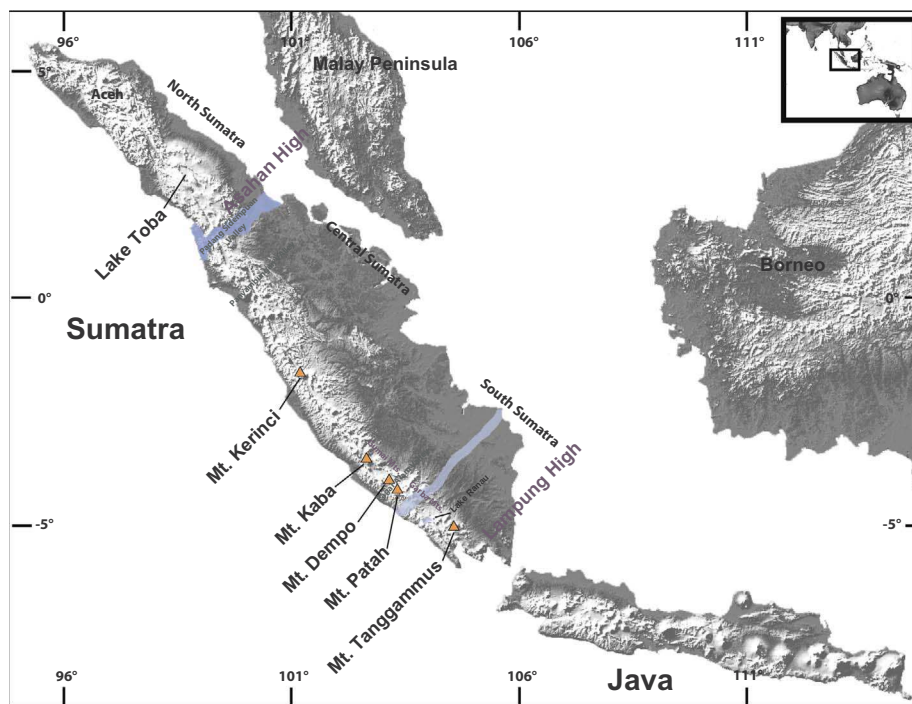


Fig. 1. Map of the islands of Sumatra and Java, showing their placement within the Sunda Shelf region. We also label historical and contemporary geological features on Sumatra referenced in this study. The two light blue waterways represent hypothesized marine incursions that may have promoted diversification of resident Sumatran taxa during the Miocene and early Pliocene.

such, Sumatra was composed of at least three large islands for much of its geologic history, and even at times of low sea levels (when marine incursions subsided), the persistence of the Padang Sidempuan and Pagar Alam Valleys likely maintained allopatric distributions of dispersal-limited species in the northern, central, and southern components of the Barisan Mountain Range (Meijaard, 2004). Equivalently, Java was composed of small volcanic islands from 10 Ma onward, and did not completely emerge above sea-level until ~5 Ma (Lohman et al., 2011). West Java may have been periodically connected to southern Sumatra via the Lampung High (elevated region similar to the Asahan High) as early as the Mid-Miocene, allowing early dispersal from southern Sumatra (van Bemmelen, 1949; Meijaard, 2004). Signals of these historical processes may be detected in the diversification histories, population structures and distributions of genetic diversity of extant biota (Weigelt et al., 2016; Portik et al., 2017; Xue & Hickerson, 2017). Under a comparative phylogeographic framework, shared diversification patterns between species can indicate synchronous responses to geological or climatological events (Hickerson et al., 2010; Bagley & Johnson, 2014; Smith et al., 2014; Prates et al., 2016).

We explore diversification processes on Sumatra and Java using species from the parachuting-frog genus *Rhacophorus*. This genus includes ~90 species distributed from the Indian peninsula to East and Southeast Asia (Frost, 2017). Sumatra and Java contain 16 described species of *Rhacophorus*, including *R. achantharrhena*, *R. barisani*, *R. bengkulensis*, *R. bifasciatus*, *R. catamitus*, *R. cyanopunctatus*, *R. indonesiensis*, *R. margaritifera*, *R. modestus*, *R. nigropalmatus*, *R. norhayatii*, *R. pardalis*, *R. poecilonotus*, *R. prominatus*, *R. pseudacutirostris*, and *R. reinwardtii* (Harvey et al., 2002; Streicher et al., 2012, 2014a,b; Hamidy & Kurniati, 2015; O'Connell et al., 2018a). On Sumatra, some species distributions span the length of the island, whereas others are restricted to small geographic areas (Harvey et al., 2002; Streicher et al., 2012; Hamidy & Kurniati, 2015). Species of the genus *Rhacophorus* occupy a variety of niche spaces, and most species' ranges are partitioned by elevation and island region (Harvey et al., 2002). On Sumatra, up to four highland endemic species occur in sympatry across the Barisan mountain range (KAO, personal observation). Java contains two species: *R. margaritifera* and *R. reinwardtii* (Streicher et al., 2012; Frost, 2017, Fig. 2).

This study uses both mitochondrial and nuclear DNA sequence data,

along with genome-wide SNPs, to pursue the following questions: (1) did species with similar geographic distributions respond synchronously to geological and climatological events on islands? (2) What historical processes promoted these diversification events?

2. Materials and methods

2.1. Sampling and molecular sequence generation

2.1.1. Taxonomic sampling

The taxonomy of several *Rhacophorus* species is currently under review; thus, we focused this study on 13 species (see discussion of *Rhacophorus* taxonomy in Appendix A). We extracted DNA from liver and thigh muscle tissue from 12 species from Sumatra and Java stored in SDS buffer or 70% ethanol. Our sampling included: *R. achantharrhena* (n = 8), *R. bengkulensis* (n = 4), *R. catamitus* (n = 27), *Rhacophorus* sp. (n = 9), *R. cyanopunctatus* (n = 3), *R. margaritifera* (n = 5), *R. modestus* (n = 23), *R. nigropalmatus* (n = 1), *R. pardalis* (n = 3), *R. poecilonotus* (n = 25), *R. prominatus* (n = 5), *R. reinwardtii* (n = 4).

2.1.2. Molecular sequence data generation and alignments

We sequenced a 609 base pair (bp) fragment of the 16S ribosomal RNA gene following O'Connell et al. (2018a). To create a multi-locus concatenated alignment, we used brain derived neurotrophic factor gene (*BDNF*) sequence data from O'Connell et al. (2018b), and downloaded sequences from GenBank of all other available *Rhacophorus* (n = 56), at least one species of each genus within the family Rhacophoridae (n = 17), eight species of Mantellidae, and two outgroups (*Rana kukunoris*, and *Occidozyga lima*) following O'Connell et al. (2018b) and Li et al. (2013). Our dataset included sequences for 12S rRNA (n = 17), 16S rRNA (n = 180), Cytochrome oxidase c subunit I (*COI*, n = 23), Cytochrome b (*CYTB*, n = 29), *BDNF* (n = 30), proopiomelanocortin (*POMC*, n = 27), recombination-activating gene 1 (*RAG1*, n = 18), Rhodopsin (*RHOD*, n = 15) and Tyrosinase (*TYR*, n = 7). All information regarding specimens used and Genbank ID is presented in Table A1 (Appendix A). We aligned each locus individually using the Geneious aligner (global alignment with free end gaps and a Cost Matrix = 65% similarity, 5.0/–4.0).

To place Sumatran and Javan species within a broad phylogenetic

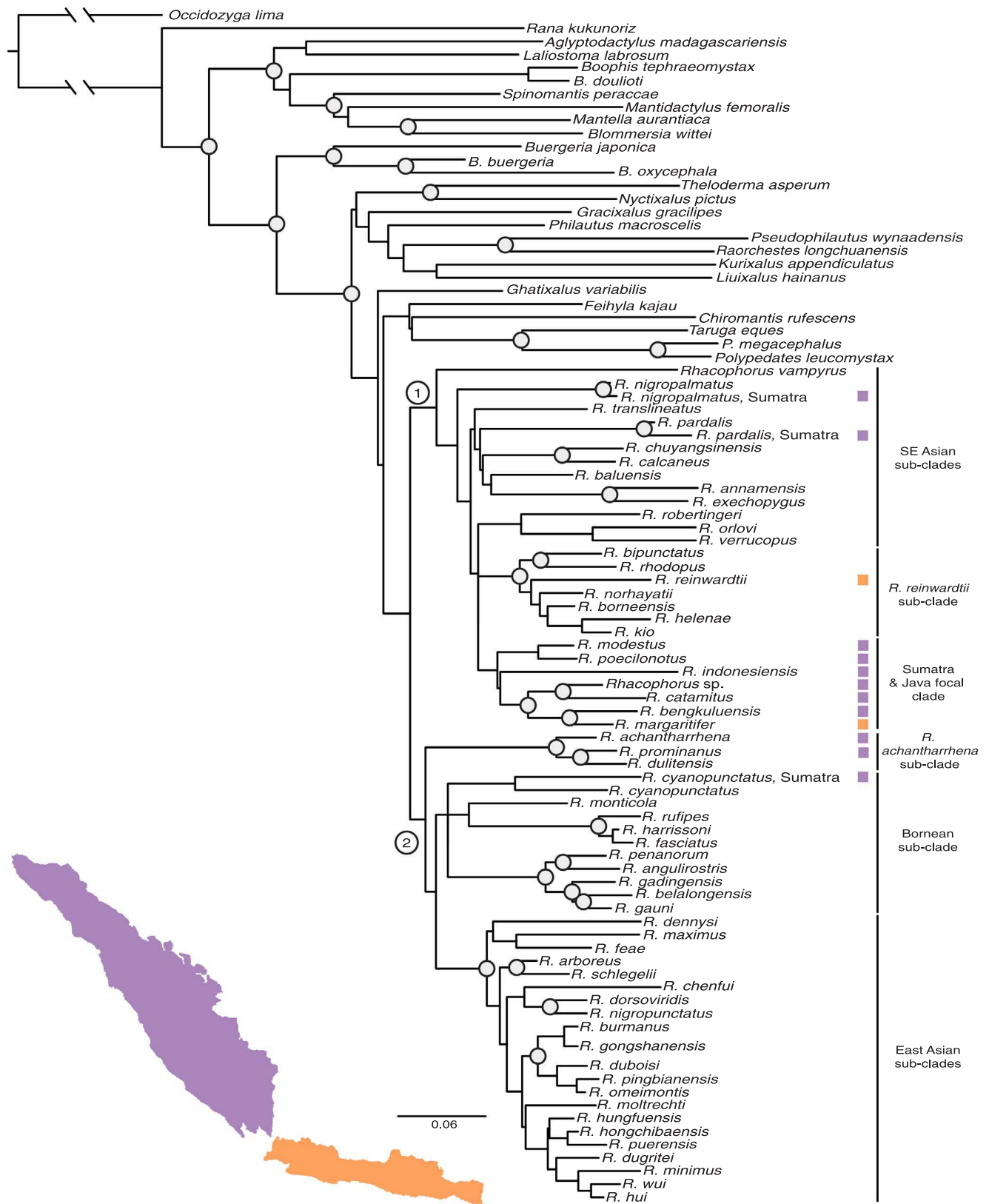


Fig. 2. Maximum likelihood phylogeny of 63 *Rhacophorus* species using concatenated mitochondrial and nuclear sequence data. Species found on Sumatra are designated with purple squares; species on Java are designated with orange squares. Subclades referenced in the study are labeled on the right. The two primary clades recovered in this study are labeled at the MRCA of each clade. Nodes with $\geq 70\%$ bootstrap support are denoted with gray circles.

context, we created an alignment that included all available loci for one individual from each *Rhacophorus* species, and included all outgroups (“phylogenetic dataset,” $n = 91$). This dataset included one to nine loci for each sample (19 *Rhacophorus* species had only 16S data). For species distributed on Sumatra and another landmass (ex. *R. pardalis*), we included a Sumatran sequence as well as a sequence from the other landmass when available. We also created an alignment that would allow us to conduct comparative phylogeographic analyses across Sumatra and Java. This alignment included all individuals from the phylogenetic dataset, as well as all 16S sequences for Sumatran and Javan species that we generated (“phylogeographic dataset,” $n = 181$).

To estimate phylogenetic relationships using additional mitochondrial loci to the exclusion of nuclear loci, we generated partial mitochondrial genomes (mtgenome) for four Sumatran species. We generated ~13,000 bp of mitochondrial sequence data for *R. achantharrhena*, *R. modestus*, *R. poecilonotus*, and *R. catamitus*. Briefly, we digested the nuclear genome using plasmid-safe DNAase, amplified the isolated mitochondrial template with whole-genome amplification, and then prepared Illumina® genomic shotgun libraries for each sample. We sequenced our libraries on a partial lane of the Illumina® MiSeq at the University of Texas Arlington Genomics Core Facility (gcf.uta.edu, Arlington, TX, USA). We filtered our raw reads using FASTX Toolkit (Gordon and Hannon, 2010), and assembled filtered contigs in CLC Genomics WorkBench (<https://www.qiagenbioinformatics.com>). We annotated each mtgenome in MitoS (Bernt et al., 2013). We downloaded additional mtgenomes of 15 species from Genbank: *Mantella madagascariensis*, *Buergeria buergeria*, *B. oxycephala*, *Gracixalus jinxiensis*, *Raorchestes longchuanensis*, *Kurixalus odontotarsus*, *K. verrucosus*, *Chiromantis vittatus*, *Polypedates braueri*, *P. megacephalus*, *Rhacophorus demysii*, *R. schlegelii*, *R. bipunctatus*, and *R. kio* (Table A1, Appendix A). We extracted coding sequences and the two ribosomal RNA sequences to create a concatenated mitochondrial alignment for phylogenetic analysis. Our alignment measured 10,837 bp in length, and comprised 12S (926 bp), 16S (1544 bp), NADH dehydrogenase subunit 1 (ND1, 746 bp), NADH dehydrogenase subunit 2 (ND2, 622 bp), COI (1488 bp), Cytochrome oxidase c subunit II (COII, 1089 bp), ATP synthase subunit 8 (ATP8, 969 bp), ATP synthase subunit 6 (ATP6, 142 bp), Cytochrome oxidase c subunit III (COIII, 858 bp), NADH dehydrogenase subunit 3 (ND3, 1315 bp), NADH dehydrogenase subunit 4L (ND4L, 250 bp), NADH dehydrogenase subunit 4 (ND4, 597 bp), and CYTB (279 bp). We call this the “mtgenome dataset.”

2.1.3. SNP data generation and processing

We prepared double-digest restriction-site associated DNA sequencing (ddRADseq) libraries for 72 individuals following Streicher et al. (2014a). We sequenced the first 60 individuals (100 bp fragments, paired end) on a partial lane of the Illumina® HiSeq 2500 at the University of Texas Southwestern Genomics Core facility (genomics.swmed.edu). To fill in sampling gaps for individuals that did not sequence well, we sequenced an additional 15 individuals on a single lane (150 bp, paired end) of the Illumina® X at the Medgenome (medgenome.com).

Double-digest RAD data were analyzed using the STACKS v1.37 pipeline (Catchen et al., 2013). After an initial round of data exploration that recovered very few SNPs, we removed all individuals with fewer than 500,000 reads, leaving 31 individuals from the original 60. No individuals were removed during processing of the second lane of sequencing. We followed the recommended workflow, which implemented the following scripts and programs: (i) process_radtags, which filtered out reads below 90% quality score threshold, (ii) ustacks, which set a maximum distance of 4 between ‘stacks,’ (iii) cstacks, which creates a catalogue of all of the ‘stacks’ within all individuals (-n flag, setting of 0) (iv) sstacks, which searches the stacks created in ustacks against the catalogue from cstacks, and (v) populations, which genotypes each individual according to the matched loci from sstacks ($-r = 0.7$). We further filtered our data with custom Python scripts

following O’Connell et al. (2017) to remove sites with more than two nucleotides and invariant sites, and to remove individuals with more than 55% missing data. This allowed us to control the amount of missing data at the locus and individual levels. We used these filtered data to create input files for downstream analyses. We analyzed each species group separately to produce three data sets, which increased the number of shared sites and minimized missing data caused by allelic dropout (Arnold et al., 2013). Our filtering retained 20 individuals and 1,755 SNPs for *Rhacophorus* sp. and *R. catamitus*, 16 individuals and 324 SNPs for *R. poecilonotus*, and 10 individuals and 504 SNPs for *R. modestus*.

2.2. Phylogenetics and comparative phylogeography

2.2.1. Phylogenetic and divergence-dating analyses

We selected the most probable model of nucleotide evolution for Bayesian inference (BI) and maximum likelihood (ML) analyses for all alignments using Bayesian information criteria implemented in PartitionFinder v.1.1.1 (Lanfear et al., 2012) partitioning by gene. The ML phylogeny for the phylogenetic dataset was estimated using raxmlGUI v1.3 (Silvestro and Michalak, 2012). Four gene partitions were defined: 12S and 16S, COI and CYTB, BDNF and RHOD, and POMC, RAG1, and TYR. We assigned a GTR + Γ rate to each partition and sampled 1,000 rapid bootstrap iterations.

Phylogeny and divergence times were estimated for Sumatran and Javan clades using the phylogeographic dataset in BEAST v.2.4.5 (Bouckaert et al., 2014). We defined three gene partitions: 12S and 16S, CYTB and COI, and all five nuclear genes. Due to a lack of run convergence using a GTR model of nucleotide evolution (ESS values < 200), we assigned the HKY model to each partition (after Drummond and Bouckaert, 2015). Following Li et al. (2013), we calibrated the origin of Rhacophoridae to 53.2 Ma based on the fossil *Indorana prasadi*. We assigned a relaxed Log Normal clock model and the constant-growth coalescent tree prior to estimate divergence times among populations within species. A Log Normal calibration was assigned to the most recent common ancestor (MRCA) of all Rhacophoridae, with a mean of 1.0, a standard deviation of 1.25, and an offset of 52.3 (the age of the fossil). This produced a Rhacophoridae MRCA distribution 95% confidence interval (CI) of 52.3–57.6 Ma. Rhacophoridae was constrained to monophyly. A uniform prior distribution was placed on the MRCA of *Boophis doulioti* and *Boophis tephraeomystax*, with a range of 0.0–15 Ma (the oldest estimated age for the Comoro island of Mayotte, where *B. tephraeomystax* is endemic (Vences et al., 2003)). An exponential distribution with a mean of 10 was assigned to the uclMean. The analysis was run for 200,000,000 MCMC generations, sampling every 20,000 generations. We checked convergence of runs (ESS values > 200), and mixing, using Tracerv1.6 (Rambaut and Suchard, 2014). We removed the first 25% of trees as burnin (2500 trees) using TreeAnnotator (Bouckaert et al., 2014), and combined the remaining 7500 trees to produce the maximum clade credibility tree with median node heights. BEAST and TreeAnnotator were run on Cipres web portal (Miller et al., 2010).

All steps of the divergence-dating analysis were repeated using the mtGenome dataset with a few modifications. Four partitions were assigned: 12S and 16S; ND2, ATP6, ATP8, and NAD4L; COI; ND4, ND1, COII, COIII, CYTB, and NAD3, with a GTR + Γ model of nucleotide evolution on all partitions. The same fossil calibration was assigned to the MRCA of Rhacophoridae, but we were unable to include the *Boophis* calibration because mtgenomes were not available for those species. We assigned a Yule tree prior.

2.2.2. SNP clustering

The program STRUCTURE (Pritchard et al., 2000) was used to explore how genomic variation is partitioned across Sumatra using SNPs. The three species groups were analyzed separately (*Rhacophorus* sp. and *R. catamitus* analyzed together) using a range of K values (1–10), with

five iterations per K value. Each analysis was run for 1,000,000 generations with a burn-in of 100,000 MCMC generations using the independent allele frequency and the admixture ancestry model. Results were summarized using the Evanno method (Evanno et al., 2005) implemented in STRUCTURE HARVESTER (Earl, 2012). We chose the highest DeltaK value and visually inspected results files at each value of K . We used CLUMPP (Jakobsson & Rosenberg, 2007) to summarize population assignments across runs and created graphical summaries using DISTRUCT (Rosenberg, 2004).

2.2.3. Estimating effective migration and genetic diversity

We visualized patterns of historical migration and the spatial distribution of genetic diversity across Sumatra using SNP data and the program EEMS (Petkova et al., 2015). EEMS estimates effective migration across the landscape by visualizing regions where genetic dissimilarity decays more quickly than expected under a model of isolation by distance (IBD). It relates effective migration rates to expected genetic dissimilarities between localities (demes) to identify barriers to migration between populations. The method also estimates genetic diversity by identifying deviations from expected genetic dissimilarity for individuals within each locality. In other words, effective diversity is a function of how differentiated individuals are within a locality, whereas effective migration is a function of how differentiated two localities are from one another. We ran three independent chains using a deme size of 500 for 8,000,000 MCMC iterations, with 3,200,000 iterations of burn-in and 9999 thinning iterations. We checked for convergence and mixing, and visualized migration and diversity surfaces using rEEMSplots in Rstudio3.1.1 (Racine, 2012; Petkova et al., 2015; R Core Team, 2016).

2.3. Evaluating diversification hypotheses

2.3.1. Testing for synchronous divergence at multiple time scales

To test for synchronous divergence at the species and population levels on Sumatra and Java, we used a hierarchical Approximate Bayesian Computation (hABC) approach as implemented by the PYMSBAYES package (Oaks, 2014a). This program can implement both msBayes (Huang et al., 2011) and dpp-msbayes (Oaks, 2014a). Both models compare the fit of empirical data to data simulated under user-informed priors. However, Oaks et al. (2013, 2014b) found that msBayes may have a bias towards supporting synchronous divergence, and corrected this bias by applying a Dirichlet process over the hyperprior specifying the number of divergence events. In response, Hickerson et al. (2014) presented modifications to msBayes using Bayesian model averaging. We used PYMSBAYES to compare dpp-msbayes and MsBayes, as well as to perform simulation-based validation for both models using 16S sequence alignments. We first estimated the number of divergence events between four sister-species pairs on Sumatra and Java: *R. catamitus* and *Rhacophorus* sp., *R. modestus* and *R. poecilonotus*, *R. bengkulensis* and *R. margaritifera*, and *R. achantharrhena* and *R. prominans*. We performed 1,000,000 simulations for each divergence model, and retained the 1000 simulations with the best fit to the empirical data to estimate posterior parameter values. We used our Bayesian divergence time estimates to guide prior selection for dppMsBayes as follows: concentration $\sim$ gamma, based on an equal prior probability of 2 or 3 divergence events between the four taxa [1000, 0.00182], tau $\sim$ gamma, based on estimates of divergence times across all taxa [1, 0.06], theta $\sim$ gamma, based on approximate mutation rates and population size [2, 0.004]. All parameters were estimated independently for each taxon-pair, and the time units were defined in substitutions per site. The msBayes analysis was run under the following priors: Theta [0.3156168, 2.666 e-08], upper limit on diverge time = 1.0, Psi = 0, upper bound of ancestral population size = 0.25. These analyses were also run without *R. achantharrhena* and *R. prominans* because *R. achantharrhena* is the sister taxon to a clade comprising both *R. prominans* and *R. dulitensis* (we lacked adequate

sampling of *R. dulitensis* to test this pair directly), and because *R. prominans* is not endemic to Sumatra (Malkmus et al., 2002; Frost, 2017).

Our second analysis estimated the number of divergence events within co-distributed high and middle-elevation Sumatran species at the oldest cladogenetic event within each species. Population pairs included central and southern *R. catamitus*, northern and central/southern *R. modestus*, central and northern/southern *R. poecilonotus*, and northern and southern *R. bengkulensis* under the same priors as above. To test for a bias towards inferring synchronous divergence we simulated 100 pseudo-observed data sets under asynchronous divergence parameters (constrained the number of divergence events to 4). We analyzed the pseudo-observed datasets under the models described above based on 500,000 simulated datasets.

2.3.2. Species tree estimation for northern, central, and southern lineages

To infer relationships between the three lineages within each species group using SNP data, we estimated species trees in SNAPPv1.3.0 implemented in BEASTv2.4.8 (Bryant et al., 2012; Kühnert et al., 2014). We limited sampling to up to four individuals per lineage within each species group to reduce computational burden, permitting the inclusion of 16 *R. catamitus*, 16 *R. poecilonotus*, and 10 *R. modestus*. We allowed BEAUTi to estimate the mutation rate for both U and V (*R. catamitus* = 1.9, 0.67; *R. poecilonotus* = 2.4, 0.63; *R. modestus* = 0.83, 1.25). We assigned a Gamma distribution to our Lambda prior, with an Alpha of 2 and a Beta of 60. This produced a median value of 100 (95% CI = 21.3–285). On our Snap prior we assigned an Alpha of 1, a Beta of 100, and a Lambda of 92.6, and placed a Gamma distribution on the rate prior. Analyses were run for 1,000,000 MCMC generations, sampling every 1000 generations. We checked convergence of runs in Tracerv1.6 (Rambaut and Suchard, 2014), visualized the complete tree sets in DENSITREEv2.2.5 (Bouckaert and Heled, 2014), and removed the first 10% of trees as burn-in.

3. Results

3.1. Phylogenetic, phylogeographic, and divergence dating analyses

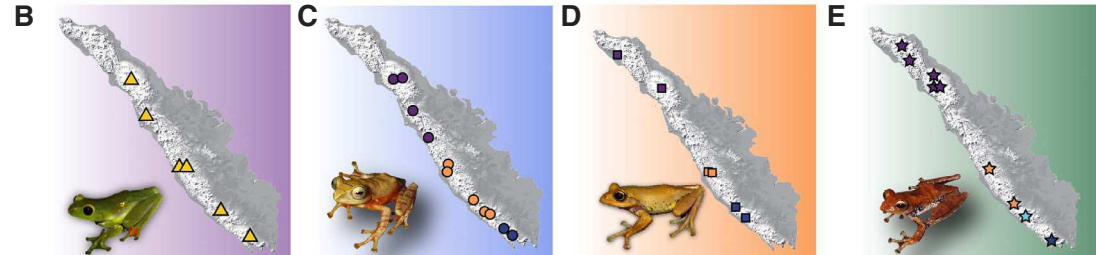
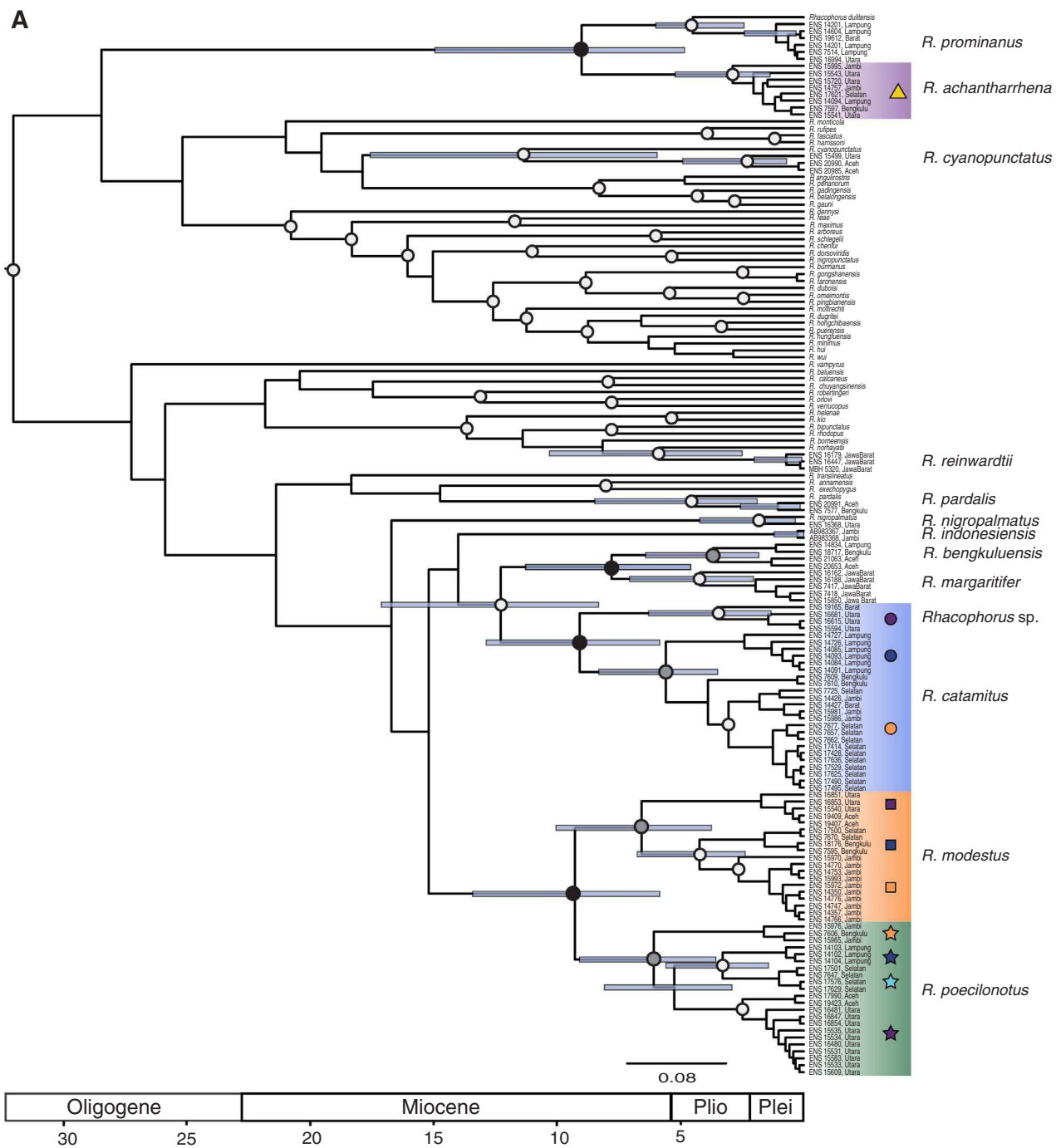
3.1.1. *Rhacophorus* on Sumatra and Java do not form a monophyletic group

Our ML analysis of the phylogenetic dataset recovered two primary clades within *Rhacophorus* (Fig. 2). Most of the deeper nodes within *Rhacophorus* were recovered with low bootstrap support (< 70%), whereas younger clades were generally well supported (Fig. 2). Sumatran and Javan species were recovered in both primary clades. Clade 1 included species from South East (SE) Asia, Borneo, Sumatra, and the Malay Peninsula grouped into four sub-clades, and included 9/14 species from Sumatra and Java (Fig. 2). Seven species endemic to Sumatra and Java formed a subclade: *R. indonesiensis*, *R. poecilonotus*, *R. modestus*, *R. catamitus*, *Rhacophorus* sp., *R. bengkulensis*, and *R. margaritifera*. In Clade 2, we recovered five sub-clades composed primarily of species from Borneo and SE and East Asia (Fig. 2). Sumatran species *R. achantharrhena*, *R. prominans*, and *R. dulitensis* formed their own subclade (*R. achantharrhena* subclade), and *R. cyanopunctatus* was most closely related to Bornean species (Fig. 2).

Our BI analysis using mtgenomes fully converged. Our phylogeny was fully resolved and recovered the two primary clades described above, with species from Sumatra and Java represented in Clade 1 by *R. catamitus*, *R. poecilonotus*, and *R. modestus*, and in Clade 2 by *R. achantharrhena* (Fig. A1). *Rhacophorus achantharrhena* was more closely related to the East Asian *Rhacophorus* than to the other Sumatran species as in the previous analysis.

3.1.2. Phylogeographic structure on Sumatra is congruent among highland species

Our Bayesian analysis of the phylogeographic dataset fully converged, and recovered a similar topology to our genus-wide analysis,



(caption on next page)

Fig. 3. (A) Bayesian phylogeny estimated using mitochondrial and nuclear sequence data of all sampled *Rhacophorus* species with a focus on population level sampling for species from Sumatra and Java. Nodes with $\geq 95\%$ posterior probability are marked by light-gray circles. Nodes for Sumatran taxa included in analyses of synchronous divergence are highlighted with larger circles. The analyses of sister-species pairs are marked by black circles (*R. achantharrhena* and *R. prominans*, *R. bengkulensis* and *R. margaritifera*, *Rhacophorus* sp. and *R. catamitus*, and *R. modestus* and *R. poecilonotus*); nodes used to test the synchrony of the oldest divergence event within co-distributed Sumatran species (*R. bengkulensis*, *R. catamitus*, *R. modestus*, and *R. poecilonotus*) are marked with dark gray circles. Species from Sumatra and Java are labeled on the right. The colored shapes correspond to the assignment of each population in focal Sumatran species shown in the maps below: purple (northern), orange (central), and blue (southern). Yellow signifies minimal phylogenetic structure. Shapes correspond to species identities: triangle = *R. achantharrhena*, circle = *R. catamitus* group, square = *R. modestus*, and star = *R. poecilonotus* group. (B–E) Maps of sampling localities and population assignments for individuals. Maps correspond to different focal species: B = *R. achantharrhena*, C = *R. catamitus* group, D = *R. modestus*, and E = *R. poecilonotus* group.

recovering Clades 1 and 2 with low support at deeper nodes (Fig. 3A). We restrict our discussion here to species on Sumatra and Java. Within these two islands, we recovered seven species that demonstrated reduced within-island lineage divergence (*R. cyanopunctatus*, *R. prominans*, *R. achantharrhena*, *R. nigropalmatus*, *R. pardalis*, *R. reinwardtii*, and *R. indonesiensis*), and six species with substantial population structure (*R. bengkulensis*, *R. margaritifera*, *R. poecilonotus*, *R. modestus*, *Rhacophorus* sp., and *R. catamitus*, Fig. 3A). We included five species with multi-landmass distributions within the Sunda Shelf, but only *R. cyanopunctatus* demonstrated deep phylogenetic structure between landmasses (note that *R. reinwardtii* and *R. prominans* were sampled only from a single landmass). Among species endemic to Sumatra or Java, all exhibited phylogenetic structure corresponding to discrete geographic regions except *R. achantharrhena* (Fig. 3A, D). Several Sumatran species shared two congruent phylogenetic breaks: a northern and central break between Lake Toba and Mount Kerinci, and a central and southern break between the Gumai and Garba Mountains (Fig. 1). For example, the boundary between *Rhacophorus* sp. and *R. catamitus* lies between Lake Toba and Mt. Kerinci, and *R. catamitus* exhibits an internal phylogenetic break between Mts. Patah and Lake Ranau (Figs. 1, 3A, E). Likewise, in *R. modestus* northern and southern lineages are divided between Lake Toba and Mt. Kerinci, and central and southern lineages are divided between Mt. Kerinci and the Gumai Mts. (Fig. 3A, F). In *R. poecilonotus*, the northern and central lineages are

divided between Lake Toba and Mt. Kerinci, and the Pagar Alam Valley divides the central and southern lineages (Fig. 3A, G). Finally, *R. bengkulensis* exhibits a phylogenetic break between lineages on mountains in Aceh province and Mts. Kaba and Tangamuss (Fig. 3A).

3.1.3. Divergence dating suggests synchrony of divergence

We used divergence dating to estimate the temporal congruence between divergence times (Figs. 3A, 4). We recovered substantial overlap in divergence times between species pairs on Sumatra and Java, including *R. catamitus* and *Rhacophorus* sp. at 9.07 Ma (5.83–12.85 Ma), *R. modestus* and *R. poecilonotus* at 9.26 Ma (5.84–13.34 Ma), *R. bengkulensis* and *R. margaritifera* at 7.77 Ma (4.59–11.30 Ma), *R. cyanopunctatus* on Sumatra and Borneo at 11.34 Ma (5.93–17.57 Ma), and *R. achantharrhena* and *R. prominans*/*R. dulitensis* at 9.0 Ma (4.83–14.92 Ma, Figs. 3A, 4). The oldest cladogenetic events within co-distributed highland and middle-elevation species were also largely congruent (Figs. 3A, 4B). In *R. catamitus* this corresponded to the divergence between central and southern lineages at 5.58 Ma (3.46–8.31 Ma). In *R. modestus* the oldest divergence event occurred between the northern and central/southern lineages at 6.56 Ma (3.77–10.10 Ma), whereas in *R. poecilonotus* it corresponded to the divergence between the central and northern/southern lineages at 6.10 Ma (3.55–9.10 Ma). Finally, in *R. bengkulensis* the oldest divergence event was between the northern and southern lineages at 3.70 Ma

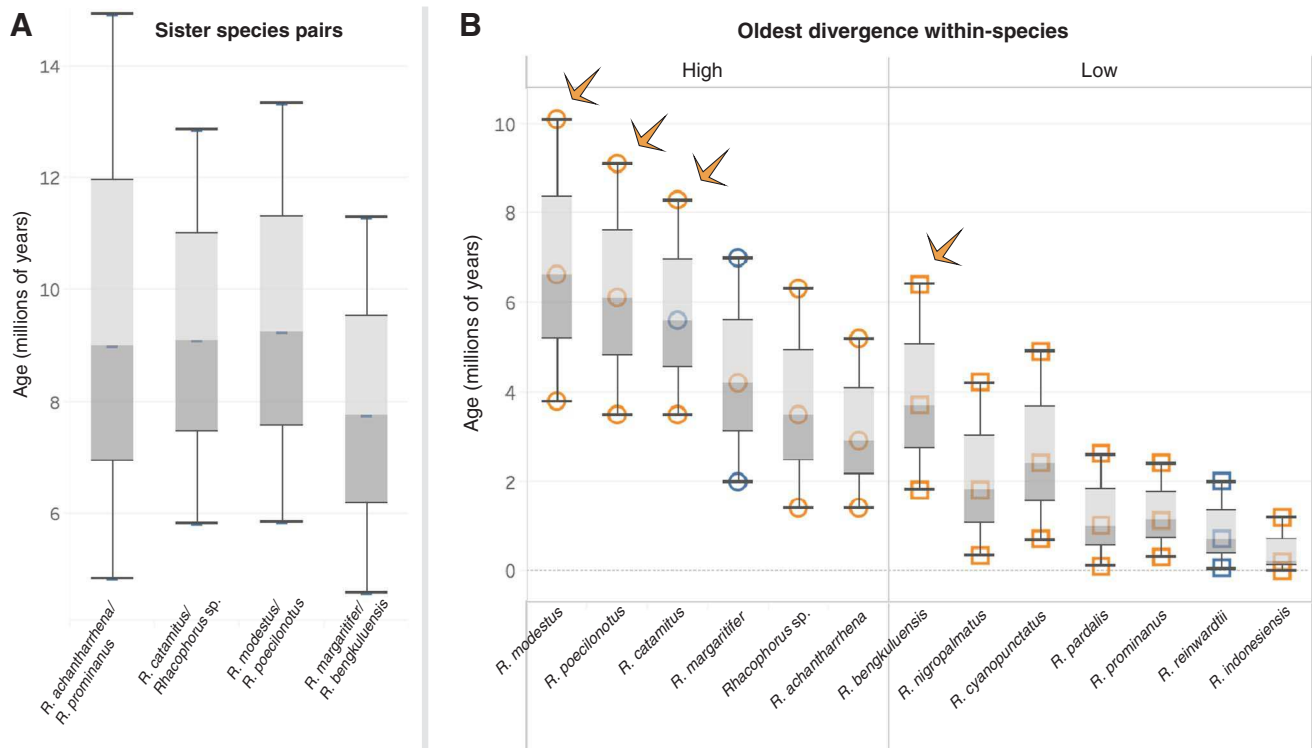


Fig. 4. Box and whisker plots showing the mean and 95% CI of divergence-date estimates from BEAST output as estimated using Sanger sequence data. (A) Divergence-date estimates of selected sister-species pairs used to test for synchronous divergence and the joint posterior probability of the average divergence time and the variance in divergence times/average divergence time. (B) Divergence-date estimates of the oldest cladogenetic event within each Sumatran and Javan species. Arrows show the four species used in the hABC analyses. Species in B are divided into high and low-elevation taxa. We defined high elevation taxa as those with ranges generally above 1000 m, and low elevation taxa as those with ranges generally below 1000 m.

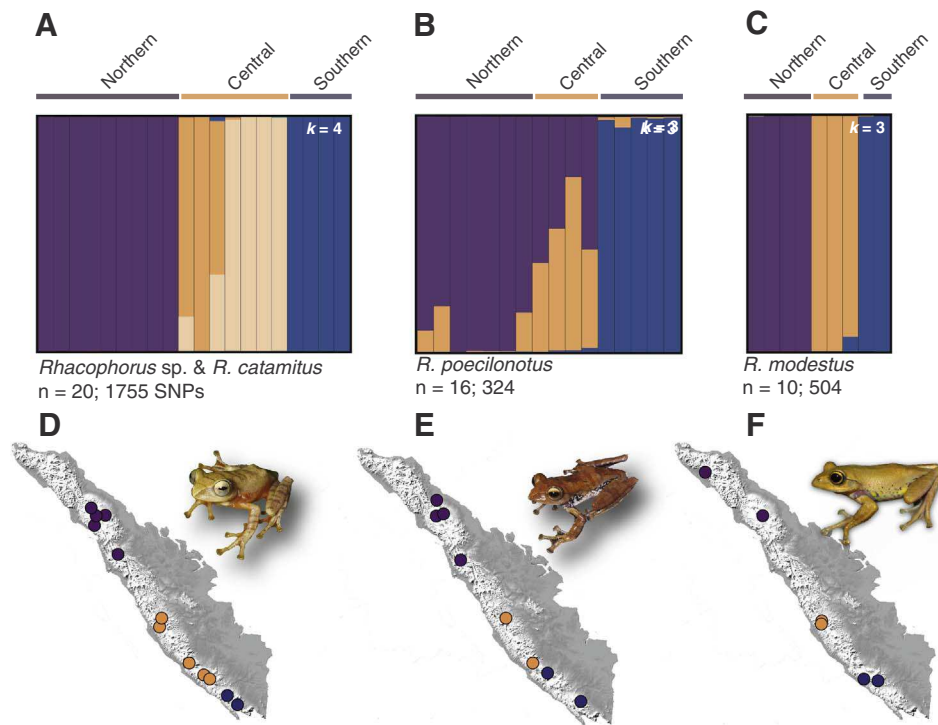


Fig. 5. (A–C) Structure plots estimated using SNP data for *Rhacophorus catamitus* group, *R. poecilonotus* group and *R. modestus*. Colors correspond to population assignments, where purple = northern, orange = central, and blue = southern. (D–F) Maps of population assignments for SNP data as inferred by STRUCTURE.

(1.81–6.38 Ma, Figs. 3A, 4B). In other Sumatran and Javan species, the oldest divergence event within the species was much younger, including *R. achantharrhena* at 2.90 Ma (1.39–5.20 Ma), *R. prominatus* at 1.12 Ma (0.31–2.41 Ma), and *R. reinwardtii* at 0.71 Ma (0.06–1.97 Ma, Figs. 3A, 4B). Lowland species were generally younger than highland species (Fig. 4B).

Divergence dating of the mtgenomes (13 mtDNA loci and no nuclear loci) revealed older divergence dates compared to the phylogeographic dataset (Fig. A1). We recovered a divergence time of 18.45 Ma (14.00–23.81 Ma) between *R. catamitus* and *R. modestus/R. poecilonotus*, and of 10.70 Ma (7.12–14.80 Ma) between *R. modestus* and *R. poecilonotus*. Within Clade 2, we found that *R. achantharrhena* diverged from the East Asian sub-clade 25.19 Ma (19.26–31.55 Ma).

3.1.4. Bayesian clustering analyses support congruent population structure on Sumatra

Our Bayesian clustering results largely recovered congruent population structure between the three species groups (Fig. 5). In the *R. catamitus* group, we recovered a northern cluster (*Rhacophorus* sp.), two central clusters, and one southern cluster (Fig. 5A, D). Similar to the mitochondrial data, we recovered a southern break corresponding to the hypothesized southern marine incursion, but the boundary between the northern and central populations lies to the south of the hypothesized northern seaway. In *R. poecilonotus* we recovered northern, central, and southern clusters, but with low population assignment probabilities for the central individuals of *R. poecilonotus* (Fig. 5B, E; $k = 2$ shown in Fig. A2). However, the boundary between the central and southern populations occurred at the Pagar Alam Valley, to the north of the hypothesized seaway. In *R. modestus*, three populations were also recovered, although our sampling was not precise enough to identify the exact geographic division between populations (Fig. 5C, F).

3.1.5. Estimates of gene flow and genetic diversity reveal two barriers to gene flow

EEMS recovered less gene flow than expected under IBD for all three species groups from SNP data (Fig. 6A–C). In each species group, we

recovered barriers to gene flow separating the northern, central and southern populations, but the location of these barriers differed slightly between species. In all three species, the discontinuity between northern and central populations occurred between the Panyambungan Valley, and Mt. Kerinci. As this discontinuity lies to the south of the hypothesized northern marine incursion, these findings suggest that this barrier may not have influenced divergence in northern Sumatra. On the other hand, the southern genetic discontinuities in *R. catamitus* and *R. modestus* correspond to the hypothesized marine barrier between modern-day Mts. Patah and Tangammas, whereas the barrier in *R. poecilonotus* lies just to the north at the Pagar Alam Valley. Estimates of effective genetic diversity were similar between species, with low diversity in the north clustered around Lake Toba, and high diversity in central and southern populations. The exception was *R. catamitus*, which had high diversity in the central population, but low diversity in the southern population.

3.2. Testing diversification hypotheses

3.2.1. Divergence was synchronous on Sumatra and Java

We found evidence in the 16S data for synchronous diversification among sister-species pairs and between co-distributed populations (Table A2, Appendix 1). Here we present the posterior probability (PP) of divergence models for dpp-msBayes, followed by msBayes. In the analysis of four sister-species pairs, we recovered the highest posterior support for synchronous divergence (PP = 0.21, 0.84) compared with two temporally distinct sets of divergence events (PP = 0.11, 0.034). The HDP interval of the dispersion index of divergence times, Ω , was 0.00–0.07, 0.00–0.03. When we analyzed only three sister-species pairs, we found stronger support for synchronous divergence (PP = 0.34, 0.89; HDP Ω = 0.00–0.05, 0.00–0.01) than for two temporally distinct divergence events (PP = 0.22, 0.04). When investigating the oldest cladogenetic event within co-distributed species, we again found the highest support for synchronous divergence (PP = 0.24, 0.84; HDP Ω = 0.00–0.07, 0.00–0.20) compared with a two-divergence scenario (PP = 0.10, 0.10). Our simulation-based validations uncovered no bias

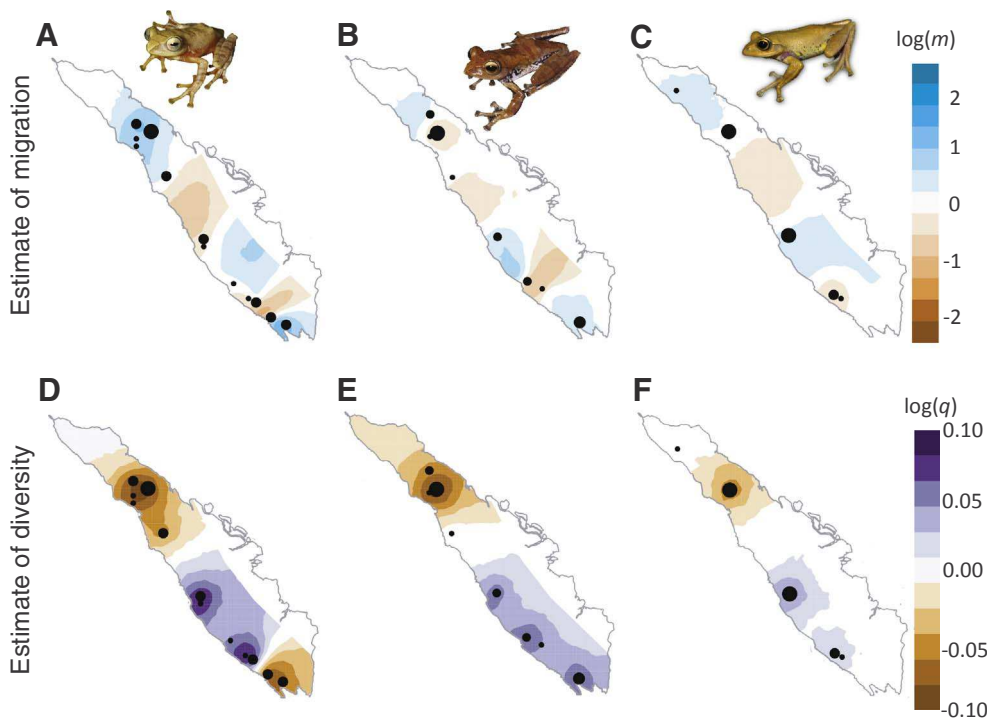


Fig. 6. Effective migration (A–C) and effective diversity surfaces (D–F) estimated using SNP data in EEMS for *Rhacophorus catamitus* group (A, D), *R. poecilonotus* group (B, E), and *R. modestus* (C, F). The size of circles represents the number of individuals in each locality (deme). In A–C, blue colors represent areas of high gene flow, or population affinity, whereas orange colors represent areas of low gene flow, or potential barriers to dispersal. In D–F, purple colors represent areas of higher-than-expected genetic diversity between two individuals within a locality; orange represents areas of lower-than-expected diversity within a locality.

towards a single divergence event in our data, although the posterior support for a single divergence event was higher in all analyses when estimated by msBayes compared with dpp-msbayes.

3.2.2. Divergence order is incongruous between highland species

The three species tree analyses from SNP data all converged and produced fully resolved tree topologies. In *R. catamitus* and *R. modestus*, the oldest cladogenetic event occurred between the northern and central-southern population lineages, whereas in *R. poecilonotus* the oldest cladogenetic event occurred between the southern and northern-central lineages (Fig. A3). These relationships corresponded to the relationships recovered in the ML analysis of sequence data for *R. catamitus* and *R. modestus*, but differed for *R. poecilonotus*, where the relationship between the three population lineages was poorly resolved but suggested that the deepest split separated the central population from the others (Fig. 3A).

4. Discussion

We used a combination of DNA sequence and SNP data to investigate the patterns and processes of diversification on Sumatra and Java. We found evidence of synchronous diversification between three Sumatran and Javan sister-species pairs, as well as between populations of four species on Sumatra. Species tree topologies suggested that *Rhacophorus* sp. and *R. modestus* originated in northern Sumatra, whereas *R. poecilonotus* and *R. catamitus* originated in central or southern Sumatra, providing evidence of synchronous allopatric diversification of two sister-species pairs. We discuss the implications of these findings and suggest some future directions for Sumatran and Javan phylogeography.

4.1. Diversification on Sumatra and Java was largely synchronous

We recovered a signal of synchronous diversification across four species pairs on Sumatra and Java, with stronger support for a single divergence event recovered by msBayes compared with PymSbayes. Although we found evidence of synchronous divergence in this subset of species, multiple temporal pulses of diversification likely produced

the full number of *Rhacophorus* species on Sumatra and Java, many of which are younger, possibly corresponding to cycles of marine incursion during the Miocene and Pliocene. The mean divergence-date estimate for sister-species pairs was 9.0 Ma (Fig. 4A). During this time in the late Miocene, sea levels were low, indicating that the two inland seaways were likely not present during this time, but the underlying valleys (Padang Sidempuan and Pagar Alam Valleys) may have served as barriers to dispersal of highland *Rhacophorus* (van Bemmelen, 1949; Haq et al., 1987; Meijaard, 2004; Lohman et al., 2011; Hall, 2012b).

We also found support for synchrony of the oldest cladogenetic event between populations within species from the Sumatran/Javan focal clade, with a mean age of 5.6 Ma (Figs. 3A, 4B). This was a time of high sea levels on Sumatra, which may have isolated the three components of the island into northern, central, and southern units (Meijaard, 2004). This was also a time of increased mountain building and subsequent volcanism (Barber et al., 2005). The correspondence of population boundaries across Sumatra with the hypothesized marine incursion in the south suggests that this barrier drove Sumatran diversification more than volcanic activity. However, we did not find support that the northern barrier served as the boundary between northern and central populations for both *R. catamitus* and *R. poecilonotus*, as individuals pertaining to the northern population were recovered to the south of the hypothesized northern barrier in both species. Nonetheless, we propose that marine incursions (with a northern barrier yet unidentified), and their underlying valleys were likely the primary barriers to dispersal on Sumatra during the Miocene and Pliocene, and promoted the synchronous divergence we observed. This hypothesis is also supported by the observation that two sister pairs (*Rhacophorus* sp. and *R. catamitus*, and *R. modestus* and *R. poecilonotus*) diversified in allopatry in northern and central/southern Sumatra (Fig. S2). More extensive geographic sampling of other taxa may help to identify the location of a shared phylogeographic break between northern and central populations.

4.2. *Rhacophorus* comparative phylogeography

Few studies have investigated phylogeographic patterns across Sumatra, instead focusing on geographically restricted species