



UNIVERSITY  
OF WOLLONGONG  
AUSTRALIA

University of Wollongong  
Research Online

---

Faculty of Science, Medicine and Health - Papers

Faculty of Science, Medicine and Health

---

2014

# Borneo and Indochina are major evolutionary hotspots for Southeast Asian biodiversity

Mark de Bruyn

*Bangor University*

Bjorn Stelbrink

*Humboldt-Universität zu Berlin*

Robert J. Morley

*Royal Holloway University of London*

Robert Hall

*Royal Holloway University of London*

Gary R. Carvalho

*Bangor University*

*See next page for additional authors*

---

## Publication Details

de Bruyn, M., Stelbrink, B., Morley, R. j., Hall, R., Carvalho, G. R., Cannon, C. H., van den Bergh, G., Meijaard, E., Metcalfe, I., Boitani, L., Maiorano, L., Shoup, R. & Von Rintelen, T. (2014). Borneo and Indochina are major evolutionary hotspots for Southeast Asian biodiversity. *Systematic Biology*, 63 (6), 879-901.

Research Online is the open access institutional repository for the University of Wollongong. For further information contact the UOW Library:  
[research-pubs@uow.edu.au](mailto:research-pubs@uow.edu.au)

---

# Borneo and Indochina are major evolutionary hotspots for Southeast Asian biodiversity

## Abstract

Tropical Southeast Asia harbors extraordinary species richness and in its entirety comprises four of the Earth's 34 biodiversity hotspots. Here, we examine the assembly of the Southeast Asian biota through time and space. We conduct meta-analyses of geological, climatic and biological (including 61 phylogenetic) datasets to test which areas have been the sources of long-term biological diversity in SE Asia, particularly in the pre-Miocene, Miocene and Plio-Pleistocene, and whether the respective biota have been dominated by in situ diversification, immigration and/or emigration, or equilibrium dynamics. We identify Borneo and Indochina, in particular, as major 'evolutionary hotspots' for a diverse range of fauna and flora. While most of the region's biodiversity is a result of both the accumulation of immigrants and in situ diversification, within-area diversification and subsequent emigration have been the predominant signals characterizing Indochina and Borneo's biota since at least the early Miocene. In contrast, colonization events are comparatively rare from younger volcanically active emergent islands such as Java, which show increased levels of immigration events. Few dispersal events were observed across the major biogeographic barrier of Wallace's Line. Accelerated efforts to conserve Borneo's flora and fauna in particular, currently housing the highest levels of Southeast Asian plant and mammal species richness, are critically required.

## Disciplines

Medicine and Health Sciences | Social and Behavioral Sciences

## Publication Details

de Bruyn, M., Stelbrink, B., Morley, R. J., Hall, R., Carvalho, G. R., Cannon, C. H., van den Bergh, G., Meijaard, E., Metcalfe, I., Boitani, L., Maiorano, L., Shoup, R. & Von Rintelen, T. (2014). Borneo and Indochina are major evolutionary hotspots for Southeast Asian biodiversity. *Systematic Biology*, 63 (6), 879-901.

## Authors

Mark de Bruyn, Bjorn Stelbrink, Robert J. Morley, Robert Hall, Gary R. Carvalho, Charles H. Cannon, Gerrit D. van den Bergh, Erik Meijaard, Ian Metcalfe, Luigi Boitani, Luigi Maiorano, Robert Shoup, and Thomas Von Rintelen

Running head: MAJOR SOUTHEAST ASIAN EVOLUTIONARY HOTSPOTS

Title: Borneo and Indochina are Major Evolutionary Hotspots for Southeast Asian Biodiversity

Authors: Mark de Bruyn<sup>1,\*a</sup>, Björn Stelbrink<sup>2,\*a</sup>, Robert J. Morley<sup>3,4,5</sup>, Robert Hall<sup>3</sup>, Gary R. Carvalho<sup>1</sup>, Charles H. Cannon<sup>6,7</sup>, Gerrit van den Bergh<sup>8</sup>, Erik Meijaard<sup>9,10,11</sup>, Ian Metcalfe<sup>12,13</sup>, Luigi Boitani<sup>14</sup>, Luigi Maiorano<sup>14</sup>, Robert Shoup<sup>15</sup>, Thomas von Rintelen<sup>2</sup>

<sup>1</sup> Environment Centre Wales, School of Biological Sciences, Bangor University, Deiniol Road, Bangor LL57 2UW, UK

<sup>2</sup> Museum für Naturkunde, Leibniz-Institut für Evolutions- und Biodiversitätsforschung an der Humboldt-Universität zu Berlin, Invalidenstr. 43, 10115 Berlin, Germany

<sup>3</sup> Southeast Asia Research Group, Department of Earth Sciences, Royal Holloway University of London, Egham Hill, Egham, TW20 0EX, UK

<sup>4</sup> Palynova Limited, 1 Mow Fen Road, Littleport, Cambs, CB6 1PY, UK

<sup>5</sup> Niko Asia Ltd, Plaza City View, Jl Kemang Timur 22, Jakarta 12510, Indonesia

<sup>6</sup> Department of Biological Sciences, Texas Tech University, Texas 79409-3131, USA

<sup>7</sup> Chinese Academy of Sciences, Xishuangbanna Tropical Botanic Garden, Yunnan 666303, P. R. China

<sup>8</sup> Centre for Archaeological Science, University of Wollongong, Wollongong, NSW 2522, Australia

<sup>9</sup> Borneo Futures Project, People and Nature Consulting International, Country Woods house 306, JL. WR Supratman, Pondok Ranji, Ciputat, Jakarta, 15412, Indonesia

<sup>10</sup> School of Archaeology & Anthropology, Building 14, Australian National University, Canberra, ACT 0200, Australia

<sup>11</sup> School of Biological Sciences, University of Queensland, St Lucia, QLD 4072, Australia

<sup>12</sup> Earth Sciences, School of Environmental and Rural Science, University of New England, Armidale, NSW 2351, Australia

<sup>13</sup> GEMOC, ARC Centre of Excellence for Core to Crust Fluid Systems, Department of Earth and Planetary Sciences, Macquarie University, NSW 2109, Australia

<sup>14</sup> Department of Biology and Biotechnologies "Charles Darwin", Sapienza Università di Roma, viale dell'Università 32, 00185 Rome, Italy

<sup>15</sup> Clastic Reservoir Systems, 10700 Richmond Avenue, Suite 325, Houston, Texas 77042, USA

*\*Joint first authors*

*<sup>a</sup>To whom correspondence should be addressed:*

Mark de Bruyn, Molecular Ecology and Fisheries Genetics Laboratory, School of Biological Sciences, Bangor University, Deiniol Road, Bangor LL57 2UW, UK. Tel: +44-1248-382344. Email: markus.debruyn@gmail.com

Björn Stelbrink, Museum für Naturkunde, Leibniz-Institut für Evolutions- und Biodiversitätsforschung an der Humboldt-Universität zu Berlin, Invalidenstr. 43, 10115 Berlin, Germany. Tel: +49-30-20938430. Email: bjoern.stelbrink@mfn-berlin.de

*Abstract.*—Tropical Southeast Asia harbors extraordinary species richness and in its entirety comprises four of the Earth’s 34 biodiversity hotspots. Here, we examine the assembly of the Southeast Asian biota through time and space. We conduct meta-analyses of geological, climatic and biological (including 61 phylogenetic) datasets to test which areas have been the sources of long-term biological diversity in SE Asia, particularly in the pre-Miocene, Miocene and Plio-Pleistocene, and whether the respective biota have been dominated by *in situ* diversification, immigration and/or emigration, or equilibrium dynamics. We identify Borneo and Indochina, in particular, as major ‘evolutionary hotspots’ for a diverse range of fauna and flora. While most of the region’s biodiversity is a result of both the accumulation of immigrants and *in situ* diversification, within-area diversification and subsequent emigration have been the predominant signals characterizing Indochina and Borneo’s biota since at least the early Miocene. In contrast, colonization events are comparatively rare from younger volcanically active emergent islands such as Java, which show increased levels of immigration events. Few dispersal events were observed across the major biogeographic barrier of Wallace’s Line. Accelerated efforts to conserve Borneo’s flora and fauna in particular, currently housing the highest levels of Southeast Asian plant and mammal species richness, are critically required. [phylogenetics; biogeography; geology; palynology; climate change; ecology]

**Text:**

Since the early research (1859-1871) of Alfred Russel Wallace, Southeast (SE) Asia has held considerable interest for the study of evolutionary biology and the field of biogeography. Megadiverse SE Asia comprises both mainland and numerous archipelagos of oceanic and continental islands (Lohman et al. 2011). The fragmented nature of SE Asia's geography is mirrored in the distribution of its extant biota and a high degree of regional and local endemism (e.g., Woodruff 2010; Wong 2011). The four biodiversity hotspots that constitute SE Asia (Indo-Burma, Sundaland, Wallacea and the Philippines; Myers et al. 2000; Woodruff 2010) are under increasing threat from the unsustainable use of native species and their habitats (Koh and Sodhi 2010; Wilcove et al. 2013).

Biotic evolution in SE Asia cannot be understood without considering the region's complex tectonic and climatic evolution (Fig. 1; Fig. S1-S4, Supplementary Information, <http://datadryad.org>, doi: 10.5061/dryad.67s40), dominated since the late Paleozoic by a series of ongoing continental collisions (Metcalf 2011a). These geological events have influenced many environmental variables e.g., geomorphology, topography, atmospheric circulation, the hydrological cycle and ocean current patterns, the opening and closure of ocean gateways, and as a consequence climate and climate change (see Fig. 2 for Quaternary sea-level scenarios), either directly or indirectly (Hall 2009; Morley 2012). Present-day SE Asia is the result of more than 300 million years of 'Colliding Worlds' (van Oosterzee 1997) characterized by continent-continent and continent-arc collisions (see Supplementary Information for reconstructions of tectonic evolution). SE Asian continental blocks (Fig. S1-S3) were all derived from the Southern Hemisphere supercontinent Gondwana, and travelled north to progressively collide and coalesce prior to the ongoing

collision with the northwards-moving Australian continent (Metcalf 2011a). By the end of the Triassic (c. 200 Ma), core Sundaland comprising Sumatra, the Thai-Malay Peninsula and most of the present-day Sunda Shelf, was part of continental Asia (Fig. S2 and S3). Vegetation during the early Cretaceous was dominated by the extinct gymnosperm family Cheirolepidiaceae, with thick cuticles suggesting a paleoclimate with some degree of water stress (Vakhrameev 1991). Angiosperms increased in number and diversity during the Late Cretaceous (Morley 2000). Occurrence of *Nypa* pollen testifies to the presence of mangroves, while common Gnetales and Laurasian conifer pollen suggest upland open woodland vegetation also subject to water stress.

Three major Cenozoic collision events had a great influence on the current archipelago setting and probably substantially affected both climate and the assembly of regional biota, possibly in tandem. The impact of the Indian-Asian collision during the Eocene (Fig. S4) on SE Asia's geography is debated but likely to have been small (Hall 2009). The terrestrial connection between India and Asia, however, had a huge effect on some SE Asian fauna (e.g., Shih et al. 2009; Klaus et al. 2010; Li et al. 2013) and flora (Morley 2000), which changed dramatically in the earliest middle Eocene (c. 49 Ma) compared to that of the Paleocene and late Mesozoic. Dispersal of many plant taxa of Indian origin into SE Asia resulted in extinction of many elements of the older Paleocene flora (Fig. S4) and by 45 Ma, pollen data indicate a diverse new flora (Morley 2000). Following the late Eocene change to a global 'Ice House' climate, early Oligocene pollen floras were of markedly lower diversity but by the end of the Oligocene most pollen types characteristic of the wet, rainforest-dominated early Miocene were in place (Supplementary Information).

The vegetation of SE Asia was similarly influenced by the Asia-Australia collision (about 25-23 Ma), although the extensive forests of the Sunda Shelf islands were not substantially invaded by Australian elements (Richardson et al. 2012). However, the collision had a major impact on the climate and vegetation of the region by disrupting the Indonesian Throughflow, the major inter-oceanic current that passes through the Malay Archipelago (*sensu* Alfred Russel Wallace). Moisture that previously moved between the Pacific Warm Pool (area enclosed by the 28.5 °C sea-surface temperature isotherm) and the Indian Ocean in the Throughflow subsequently fell on Sundaland (Morley 2006). This climate change coincided with the development of the East Asian Monsoon across China (Sun and Wang 2005). The Australian-Sunda collision thus appears to have initiated the East Asian monsoon (Morley 2012). This change to a wetter climate led to the development of the modern-aspect Malesian flora (Morley 2000).

The third collision, of the westernmost Australian promontory the Sula Spur with Sulawesi during the Miocene, led to the first amalgamation of continental fragments derived from Sundaland and Australia, respectively (Hall 2009, 2011). Emergence of land in central and eastern parts of Sulawesi (Hall 2009, 2011) is supported by biological evidence such as timing of the first major colonization wave of Sulawesi by terrestrial animal taxa from both east and west of Wallacea (the group of islands lying between the Asian and Australian continental shelves, including Sulawesi, the Moluccas and the Lesser Sunda islands) in the late Miocene and Plio-Pleistocene (Sahul and Sunda Shelf; Stelbrink et al. 2012). Several islands (e.g., Borneo, Sumatra, Java) west of ‘Wallace’s Line’ (the major biogeographic barrier following the Indonesian Throughflow, running between Borneo and Bali to the west, and Sulawesi and Lombok to the east, i.e. the Makassar and



Lombok Straits) are largely of continental origin, being linked to the Sunda Shelf, and formed a contiguous landmass when sea-level dropped (Sundaland). The Sunda Shelf attained its greatest land area during the middle Eocene (c. 49-45 Ma) with predominantly southward flowing rivers and widespread, palm-dominated (Harley and Morley 1995) peat swamps across what is now southeastern Borneo, Java and southern Sulawesi. An easternmost rift of Sundaland resulted in the middle Eocene isolation of SW Sulawesi from mainland Borneo, forming the Makassar Strait and separating Sulawesi from Sundaland (c. 45 Ma; Hall 2009), creating the central section of the major biogeographic break today recognized as 'Wallace's Line'. From the later middle Eocene (c. 42 Ma) onward, southern Sundaland subsided, and by the end of the Oligocene (c. 25 Ma) much of the region now occupied by Java and southern Borneo (Kalimantan) was submerged, characterized by widespread shallow shelves with many extensive reefs (Fig. 1, 25 Ma). However, central Borneo and its northwestern extension to the Asian mainland remained emergent throughout this time (Fig. 1, 25 & 20 Ma) enabling everwet rainforest expansion northwards c. 25 Ma (e.g., Morley 2000; see below).

The onset of wetter conditions in the latest Oligocene/Miocene, and Plio-Pleistocene climate-driven sea-level changes (Fig. 2) are superimposed upon this dynamic geological framework, leading to the repeated inundation of the region's continental shelves during interglacials (Hanebuth et al. 2011). During the early Quaternary, both Sundaland and the Sahul Shelf (Australian continental shelf) underwent fundamental changes in character as the amplitude of sea-level fluctuations increased (Zachos et al. 2001), resulting in these previously submerged shelves being exposed and flooded during glacials and interglacials, respectively (Fig. 2) (Voris 2000; Hanebuth et al. 2011). This phenomenon has been best

studied across Sundaland, which over the past ~2 million years effectively doubled in size during glacials, and currently exhibits its smallest geographical area for that period with biota generally in a state of refuge (Cannon et al. 2009; Woodruff and Turner 2009). For most of the Quaternary sea levels fluctuated ~40-60m below present-day levels, and climate was neither so dry nor cool as during the Last Glacial Maximum (LGM). The dual impact of geological and global climatic cycles have generally been regarded as important factors contributing to the region's biotic assembly (Hall 2009; Lohman et al. 2011).

Building on earlier work on SE Asian biogeography (e.g., Wallace 1869; Mayr 1944; Simpson 1977; Whitmore 1981, 1987), recently refined geological and tectonic models from this region provide information on the timing and distribution of various landmasses and islands (Hall 2009, 2011, 2012a) (Fig. 1, Supplementary Information). These models underpin biological models of historical distribution of ecosystems through time (Fig. 2) informed by paleontological and palynological data and based on paleoclimatic interpretations (e.g., Cannon et al. 2009). Recent studies provide estimates of divergence times, vicariance and dispersal events and can, in combination with other empirical geological and biological data, identify key areas for the generation and maintenance of biotic diversity (e.g., Stelbrink et al. 2012). However, the complex geological history of SE Asia raises issues in the testing and interpretation of single-taxon studies, particularly across the region's heterogeneous mix of volant vs. non-volant terrestrial and freshwater taxa.

Thus, we here conduct meta-analyses and syntheses of geological, climatic and biotic (including 61 phylogenetic) datasets from the SE Asian region to examine the assembly of the region's ecosystems and biota through time and space. We examine

whether the SE Asian biota has been characterized primarily by *in situ* diversification, immigration and/or emigration, or equilibrium dynamics between our pre-defined biogeographic areas across three major time-periods for which phylogenetic and paleogeographic data were available, the pre-Miocene, Miocene and Plio-Pleistocene. We use divergence time and ancestral area estimation to test whether the sources of long-term biological diversity in SE Asia accord with predictions based on our geological, climatic and biotic meta-analyses and syntheses. These predictions, based on major events in SE Asia's known history presented above, include: 1. Levels of *in situ* diversification events – i.e. the accumulation of new lineages within areas – should reflect emergent ages and sizes of our pre-defined areas; 2. Areas with recent histories of sea-level inundation (e.g., Sumatra and Java) should be characterized by higher levels of (recent) immigration than emigration; 3. Number of colonization events should be greater between adjacent areas compared to more distant areas; 4. An increase in diversification rates should be evident during the Plio-Pleistocene resulting from increased connection-disconnection events related to sea-level fluctuations.

## MATERIALS AND METHODS

### *Southeast Asian Geological Reconstructions*

Paleogeographic reconstructions for the Paleozoic and Mesozoic follow those published in Metcalfe (2011b). Maps were generated using a wide range of multidisciplinary data including stratigraphic, sedimentological, biostratigraphic, biogeographic, paleomagnetic, paleoclimatic, structural, isotopic and geochronological data and plutonic and volcanic activity (e.g., Wakita and Metcalfe 2005). Tectonic reconstructions for the late Jurassic to

present depicting the evolution of continental blocks, ocean basins and distributions of land and sea are modified versions of Spakman and Hall (2010) and Hall (2012a) and were produced using geodynamic reconstruction modeling software. The methodology and software used is described in Hall (2002, 2012a). The distribution of land and sea is shown for a larger area than in previous publications, extending further north, as discussed in (Hall 2012b, 2013) and includes additional details of Sunda shelf marine embayments (Morley and Morley 2013) and shows freshwater rift lakes and inland seas for the Sunda Shelf (Shoup et al. 2012).

#### *Development of the SE Asian Flora Inferred from Palynology*

For the Cenozoic, patterns of paleoclimate change have been established from palynological records generated by petroleum industry studies (partly unpublished) from across the region (Morley 2012). These records total in excess of 150 datasets, extending from East and West Java Seas, Makassar Straits, offshore Sabah and Sarawak, Nam Con Son and Cuu Long Basins offshore Vietnam, the West Natuna Basin in Indonesia, the Malay and Penyu Basins offshore West Malaysia and Gulf of Thailand Basins, the more important of which are published (e.g., Morley and Morley 2011). All the datasets are placed within a high-resolution sequence biostratigraphic framework (following Morley et al. 2012) allowing paleoclimates to be consistently reconstructed for individual time slices across the entire region. The high-resolution biostratigraphy used to provide the temporal framework for the Sunda Shelf paleogeographic maps was undertaken by characterising transgressive-regressive cycles using a sequence biostratigraphic approach initially outlined in Morley (1996). The detailed stratigraphic framework used for preparing the Sunda shelf

maps is presented in Morley et al. (2012). The time slices for the Sunda Shelf were mapped seismically, interpreted in terms of paleoenvironment and presented in Shoup et al. (2012). Those maps were then simplified, rescaled and incorporated into the regional paleogeographic maps of Hall (2011a). For Sumatra, basin outlines and stratigraphy was obtained from Moulds (1989) and Aldrich et al. (1995), with environments interpreted from unpublished data. For the Quaternary, maps have been prepared using published palynological and other climate-indicative datasets combined with climate modeling techniques (e.g., as in Cannon et al. 2009 and biogeographic considerations, as in Morley 2012).

#### *Phylogenetics: Molecular Clocks and Ancestral Area Estimation*

A literature search was performed for published molecular phylogenies including taxa (subsequently assigned to the following categories: plants, insects & spiders, freshwater crustaceans, freshwater molluscs, freshwater fishes, herpetofauna, birds, and mammals) from SE Asia (and adjacent regions). Datasets were not included *a priori* if, firstly, sample sites were restricted to less than three areas pre-defined for the ancestral area estimation, and secondly, sequences, fossil data for calibration/substitution rates, and outgroup taxa were not available for performing molecular clock analyses. Some potential biases/artifacts might occur while performing meta-analyses such as that conducted here: 1) sampling artifacts, i.e., weak taxonomic and geographic coverage due to sampling area access problems; 2) incomparable divergence times among intraspecific and interspecific datasets. However, we screened datasets carefully and omitted a substantial number ( $n = 68$ ; see Supplementary Information for a list of excluded publications/datasets) prior to further

analyses to reduce potential sampling bias (see Table 1 and 2 for details). Sixty-one datasets were retained for the final molecular clock analyses and ancestral area estimation, the majority ( $n=45$ ) of which were performed or re-run for the present study using taxon-specific calibration points or substitution rates from several source publications (Table 1).

Sequences were aligned using MAFFT (default settings; <http://www.ebi.ac.uk/Tools/msa/mafft>; Katoh and Toh 2008)) and corrected by eye if necessary, i.e., removing potential gap artifacts produced by MAFFT (see Supplementary Information for alignment files). Alignments were reduced to unique haplotypes using DAMBE v. 5.1.1 (Xia and Xie 2001). Best-fitting substitution models were estimated using jModelTest v. 0.1.1 (Posada 2008; 24 models, AIC; Table 2). Substitution rates or fossil calibration points were obtained from the publication or secondary sources (Table 2). Phylogenetic dating analyses were conducted with BEAST v. 1.6.2 (Drummond and Rambaut 2007) using an uncorrelated lognormal relaxed-clock model. MCMC chains were run using a Yule tree prior for 20 million generations, or 40 million generations if ESS was  $<200$  for several parameters. Fossil, geological and indirect calibration points (see Table 2) were assigned normal distributions. Maximum clade credibility (MCC) trees were summarized using TreeAnnotator (BEAST package; burnin=35,001). All BEAST log files were visualized and checked in Tracer v. 1.5 (Rambaut and Drummond 2007) to ensure ESS values were  $>200$ . In a few cases ( $n=8$ ), ESS values did not reach values  $>200$  for ‘prior’ and ‘posterior’ while the remaining parameters show ESS values considerably higher than 200. In those cases, we performed re-analyses using the less complex HKY model (plus Gamma and PropInv parameters when selected by jModelTest for the best-fitting model), because low ESS values might indicate over-parameterized substitution

models (see e.g., Grummer et al. 2014). Importantly, divergence time estimates did not differ substantially between default and modified analyses for the vast majority of datasets (cf. Table 2 and Table S1 and S2). Distribution of relaxed-clock (BEAST) divergence times (credibility intervals and mean ages) are shown in Fig. S5-S13; see Supplementary Information for BEAST input xml files.

Localities of specimens were obtained either from the source publication or from web sources (e.g., GenBank) if needed and were assigned to one of 12 areas of endemism, partly modified from areas of endemism proposed by Turner et al. (2001): Indochina = Indo-Burma *sensu* Myers et al. 2000, Thai-Malay Peninsula, Sumatra, Philippines, Palawan, Borneo, Java, ‘Wallacea excl. Sulawesi’, Sulawesi, ‘East of Wallacea’, India and Sri Lanka, and Japan, or to additional areas such as ‘Africa’. A presence-absence matrix was created for the ancestral area estimation conducted using Lagrange (Lagrange configurator: <http://www.reelab.net/lagrange/configurator/index>) (Ree et al. 2005; Ree and Smith 2008); see Supplementary Information for Lagrange input and output files). Analyses were performed with default (unconstrained) settings (i.e., equal dispersal rates, single time matrices) and were constrained to a maximum range size of 2 areas because for the vast majority of datasets specimens could be assigned to single pre-defined areas only (see Supplementary Information). Increasing the number of ancestral ranges allows for widespread ancestors, however, by limiting this number the determination of dispersal directionalities across two areas is simplified.

Lagrange results (i.e., estimated ancestral areas for each node with the highest relative probability) were transferred to BEAST MCC tree printouts for each dataset. Changes in geographic range representing colonizations from one area to the other have

been included for resolved branches (posterior probabilities >0.5), though a higher threshold would increase the reliability of performed ancestral area estimation (see Supplementary Information for BEAST tree files). BEAST mean ages plus ancestral areas of ancestral and descendent nodes were used to assign these colonization routes to the pre-selected time frames ‘pre-Miocene’, Miocene, and ‘Plio-Pleistocene’ (until the present) (Fig. 3 and Fig. 4; see Fig. S14 and S15 for group-specific colonization routes through time among taxonomic groups). Counting dispersal events within each time bin allows for comparison of time-frame specific patterns with paleogeographic reconstructions for that period (Fig. 1).

Speciation and dispersal through time were explored by counting and analyzing number of lineages, *in situ* (intra-area) diversification, and emigration events in each area across all taxonomic groups (Fig. 4 and 5, Table 3, Table S3 and S4). The results were subsequently subjected to pairwise rank-based Mann-Whitney *U* tests using SPSS Statistics v. 17.0.0 (SPSS Inc.) to test the null hypothesis of equality between areas (asymptotic significance, two-tailed; Supplementary Information).

#### *Mammal and Plant Species Richness Across SE Asia*

Comprehensive distribution data exists for SE Asian plants and mammals, allowing comparison of extant species richness of these groups across our pre-defined areas to that recovered from our phylogenetic and ancestral area analyses. To construct a map of mammal species richness, we collected information on the current distribution of all mammal species occurring in the study area through the Global Mammal Assessment (Rondinini et al. 2011). Considering an area ranging from Myanmar to New Guinea, we



covered a total of 1,086 mammal species, belonging to 17 orders (Catullo et al. 2008); Rodentia (number of species = 388), Chiroptera (328), Primates (76), Diprotodontia (58), Soricomorpha (56), Carnivora (54), Artiodactyla (52), Scandentia (17), Dasyuromorphia (16), Peramelemorphia (12), Erinaceomorpha (7), Lagomorpha (6), Perissodactyla (3), Pholidota (3), Dermoptera (2), and Proboscidea (1). For each species we obtained the most updated available global distribution range. We weighted each species according to the inverse of its area of distribution (i.e., species presence is divided by the square kilometers of their distribution ranges) and calculated a map of species richness (10 arcminutes spatial resolution) (Fig. 6). Endemic species and species mostly abundant in Southeast Asia, ‘locally restricted taxa’ with a high number of specimens in a comparatively small area, will result in a higher weighted value and are thus considered more important in the analyses.

Data for vascular plants was used from Kreft and Jetz (2007) to create a map (Fig. 6) of species richness, with permission (*Copyright (2012) National Academy of Sciences, U.S.A.*). Considering 1,032 geographic units worldwide and a set of bioclimatic and environmental variables, the source authors generated a set of predictions for species richness (1 degree resolution) (Kreft and Jetz 2007). Among all models of species richness available from Kreft and Jetz (2007), we considered the kriging estimate here as it was demonstrated that the inclusion of neighborhood effects substantially improves the quality of predictions.

## RESULTS

*Setting the Scene: Palynological and Geological Syntheses Suggest Major Wet Refugia in Indochina and Borneo*

The early Miocene collision of Australia with the eastern margin of Sundaland dramatically changed the region's paleogeography (Hall 2009). This collision resulted in the formation of widespread uplands in central Borneo from the early Miocene onward, and the development of major rivers such as the Mahakam and Baram and their respective deltas. The Meratus Mountains (SE Borneo) were uplifted during the middle and late Miocene (Fig. 1, 15 and 10 Ma) (Witts et al. 2012) and Mount Kinabalu during the late Miocene (Hall 2011). Widespread evergreen rainforests would have covered much of Sundaland during the early and middle Miocene. There was some climate variability with drier climates being widespread during periods of low sea level at subequatorial latitudes, and with rain forests reaching their maximum northward extent during the middle Miocene. However, whereas palynological data from the late Miocene and Pliocene suggest periodicity of climate in the region of the Sunda Shelf to the west, high resolution palynological data (Supplementary Information) back to at least 9 Ma from East Borneo suggest continued everwet climates characterizing periods of both high and low (glacial) sea level in the equatorial region (Morley and Morley 2011). This climatic stability has been invoked as an explanation for high floristic diversity witnessed in areas such as Lambir Hills in Sarawak (Borneo), which harbors rain forests rivaling those of the Neotropics. The northern Sunda Shelf was subject to extensive marine inundation, with the development of widespread inland shallow seas reaching their greatest extent by the middle Miocene thermal (and sea-level) maximum (Fig. 1, 15 Ma).

### *Phylogenetic Meta-Analyses: Origins of SE Asian Biota*

Our meta-analyses of published molecular datasets using a standardized relaxed Bayesian dating approach (Supplementary Information) identified a history of phylogenetic lineage diversification in SE Asian biota since the Jurassic (Fig. S5-S13, Table S1 for detailed results). Only two of our study groups date back to the Mesozoic (plants, insects & spiders). While SE Asia's continental core – Indochina and Sundaland – was essentially established by the end of the Mesozoic, its current largely insular topography and megadiverse biota were decisively shaped by Cenozoic events. For the vast majority of phylogenetic datasets examined, diversification events were Cenozoic in age (Fig. S5-S13, Table S1 for detailed results). Geographic and biotic changes are examined here in light of the interplay of tectonics and climatic oscillations, most notably evident during three distinctive stages in the Paleogene (~65.5-23 Ma), Neogene (~23-2.6 Ma), and the Quaternary (~2.6 Ma-present; see timeline in Fig. S13).

### *Phylogenetic and Ancestral Area Estimation: Borneo and Indochina – Major 'Evolutionary Hotspots'*

Diversity can be evaluated here as the number and variety of lineages found in a specific pre-defined area (Table 1). This diversity may have arisen through the accumulation of immigrants, by *in situ* diversification, or some combination of the two. Indochina, the Philippines and Borneo held the highest total number of lineages across all taxonomic groups ( $n > 300$ ; Fig. 5a), which were statistically significantly higher than the other pre-defined areas: 'East of Wallacea' ( $n=295$ ), the Thai-Malay Peninsula ( $n=233$ ), Sulawesi ( $n=199$ ), Sumatra ( $n=168$ ), Java ( $n=93$ ), 'Wallacea excl. Sulawesi' ( $n=74$ ), and Palawan

( $n=46$ ) (Mann-Whitney  $U$  test:  $p<0.05$ ; Table S4); other significant differences in pairwise comparisons are due to low numbers. These differences remained when population-based datasets were excluded (Fig. 5a, Table S4).

Significantly higher numbers of *in situ* diversification events were observed in several areas: Indochina compared with all areas except ‘East of Wallacea’ (see Fig. 5b;  $p<0.05$ ; Table S4); Borneo compared with Sumatra, Palawan, Java and ‘Wallacea excl. Sulawesi’ ( $p<0.05$ ; however, this can primarily be attributed to particular taxonomic groups, namely plants, fishes, herpetofauna, and birds); and Sulawesi compared with Palawan and ‘Wallacea excl. Sulawesi’ for most taxon groups ( $p<0.05$ ). However, inclusion of population-based datasets (see Table 1) increased both the number of total lineages and diversification events. The removal of population-based datasets reduced significant differences to some extent, particularly for Indochina (Fig. 5; Table S4).

Indices for total number of lineages and *in situ* diversification events for each pre-defined area (excl. population-based datasets) were calculated in an attempt to control for potential biases resulting from increased lineage sampling for a particular area. These indices were tested for significant differences between areas (Mann-Whitney  $U$  tests, Table S4). A significantly higher proportion of *in situ* diversification events were still observed for Indochina (*vs.* Thai-Malay Peninsula, Sumatra, Palawan, Java, and ‘Wallacea excl. Sulawesi’;  $p<0.05$ ) and Borneo (*vs.* Java;  $p<0.05$ ), when considering total number of lineages sampled per area (excl. population-based datasets).

Because statistically significant higher total lineage numbers and number of *in situ* diversification events for Indochina and Borneo could simply have been a function of area size, an area size-dependent index for all taxonomic groups among all areas was calculated

for these parameters (Table 4). Interestingly, the comparatively large current size of both Indochina and Borneo was not reflected by higher indices reflecting size-dependent diversity (total number of lineages and number of *in situ* diversification events). A comparatively lower diversity-size index was found for the biogeographic area ‘East of Wallacea’, likely related to under-sampling of Sahul Shelf representatives. In contrast, Sulawesi and the Philippines – both biogeographic areas with comparatively high numbers of lineages and *in situ* diversification events – showed comparatively high indices for their relatively small geographic sizes. Similarly, Palawan, the smallest area examined here showed the highest indices for these two parameters (Table 4). One important caveat on these findings is that current area size is not necessarily representative of the extended timeframe under investigation here.

Indochina and Borneo have been major colonization sources reflected by a significantly higher number of emigration events through all time periods examined here compared with our other pre-defined areas (see Fig. 4; Mann-Whitney *U* test:  $p < 0.01$  for Indochina and  $p < 0.05$  for Borneo; Table S4). Thirty-two of 49 datasets analyzed for colonization routes show ‘Out-of-Borneo’ signatures (total number of emigrations:  $n=89$ ; total number of emigrations excl. India/Sri Lanka and Japan:  $n=87$ ; compare with Fig. 4 and Fig. S14 and S15), while 36 datasets show colonizations out of Indochina (total number of emigrations:  $n=95$ ; total number of emigrations excl. India/Sri Lanka and Japan:  $n=73$ ; Table 3 and Table S3). An increased total number of emigrations were observed in the Miocene, and even more pronounced in the Plio-Pleistocene (pre-Miocene:  $n=30$ , Miocene:  $n=122$ , Plio-Pleistocene:  $n=169$ ), the latter significantly different to that of the pre-Miocene (Mann-Whitney *U* test:  $p < 0.01$ ; Table S4). In contrast, immigration events are comparably

much less frequent for these two areas (Indochina:  $p < 0.01$ , Borneo:  $p < 0.05$ ), with other areas showing a relatively balanced or even negative emigration-immigration index, e.g. Palawan and Java (Fig. 4, lower panel; not significant except for Java  $p < 0.05$ ; Table S4). Interestingly, and as expected, very few emigration events were observed from younger volcanically active emergent areas such as Java ( $n=7$ ) (Fig. 4, Fig. S14 and S15). Despite their geographical proximity, colonizations of Sulawesi from Borneo have been infrequent ( $n=6$ ; Fig. 4, Fig. S14 and S15). The total number of immigration events found across all areas is significantly different between the three timeframes, increasing towards the present (pre-Miocene:  $n=26$ ; Miocene:  $n=86$ ; Plio-Pleistocene:  $n=149$ ; pre-Miocene vs. Miocene:  $p < 0.01$ ; pre-Miocene vs. Plio-Pleistocene:  $p < 0.001$ ; Miocene vs. Plio-Pleistocene:  $p < 0.05$ ; Table S4).

A pairwise comparison of the number of datasets observed for a particular source-destination route again highlighted Borneo and Indochina as predominantly ‘sources’ rather than ‘destinations’ across SE Asia through time, e.g.: Borneo–Thai-Malay Peninsula:  $21 \leftrightarrow 5$ , Borneo–Philippines:  $14 \leftrightarrow 3$ , and Indochina–Thai-Malay Peninsula:  $15 \leftrightarrow 5$  (source  $\leftrightarrow$  destination number of datasets; Table 3). Further, the rate of colonizations between currently adjacent islands/landmasses is generally higher compared to between currently remote areas (Table 3; see also Fig. 4, Fig. S14 and S15). However, even ‘East of Wallacea’ appears to have been colonized to a considerable extent from Indochina and Borneo, even though geographic distances separating these areas have remained comparatively large throughout the timeframe under investigation (Fig. 4). The majority of colonization events were assigned to the Plio-Pleistocene, and most frequently between areas formerly adjoined via ‘Sundaland’ (Fig. 4, Table 3). To control for total number of

lineages sampled per area (sampling effect), we calculated an index for emigrations and total lineages (excl. population-based datasets) (Table S4). These indices were tested for significance between areas (Mann-Whitney  $U$  tests). Interestingly, significant differences were only found between Sulawesi vs. Indochina, Sumatra, and Borneo (lower indices for Sulawesi); and Borneo vs. Java (lower indices for Java; Table S4). This result indicates that the significantly higher levels of emigration events identified for Indochina and Borneo are clearly related to the higher diversity (represented by number of lineages) found in these areas. Elevated levels of emigration events found in Indochina and Borneo might actually relate to the size of the respective source area. However, size-dependent indices were not comparatively elevated for either of these biogeographic areas (Table S4), although current area size was used for these analyses and area size has changed considerably through time for several areas (Fig. 1).

Performing multiple pairwise statistical tests can include amendments such as the (sequential) Bonferroni correction, i.e., effectively dividing the p-value cutoff by the number of tests performed or by identifying rank-based Bonferroni-corrected p-value cutoffs. For the majority of pairwise comparisons made (e.g., number of lineages, in situ diversifications, emigrations) the number of tests was  $n=45$ . Applying a Bonferroni correction resulted in a decreased p-value cutoff of 0.0011 and therefore a greatly reduced number of significant differences observed (see Table S4). However, several drawbacks have been identified for such corrections (see e.g., Perneger 1998; Moran 2003; Armstrong 2014), thus, here we base our interpretations on non-corrected significance testing – effectively prioritising each individual area-pair test (see e.g., Armstrong 2014).

### *Mammal and Plant Distributional Datasets: Elevated Diversity on Borneo*

A meta-analysis of mammalian and plant distributional datasets for which high-resolution data across SE Asia was available (Fig. 6), again identified Borneo in particular as an area of special biodiversity significance for SE Asia (Fig. 6). For our pre-defined areas, Borneo hosts the highest levels of weighted (see Material and Methods) terrestrial mammalian ( $n=230$ ) and vascular plant species richness in SE Asia, while the top 10% of weighted mammalian species richness is hosted in the northeast of the island and in the Thai-Malay Peninsula ( $n=228$ ; Fig. 6). This pattern was consistent for Borneo for total number of plant lineages identified here (Fig. 5a and Table S4) but was not the case for mammals. For mammalian lineages, Borneo ranked fourth after Indochina, the Philippines and the Thai-Malay Peninsula. Interestingly, the Philippines showed a marked contrast between present-day mammal species richness (comparatively low) and total number of mammalian lineages (comparatively high) (cf. Fig. 5a and Fig. 6). Size-dependent indices of general diversity across all taxonomic groups were additionally applied to present-day mammal species richness (Fig. 6). However, when *current* island size is taken into account, mammal diversity (total number of species) is comparatively low for Indochina and Borneo, while it is comparatively high for Palawan, the Thai-Malay Peninsula, the Philippines and Java (Table S4).

### *Palynological and Paleontological Syntheses: Middle to Late Quaternary Sunda Shelf Sea-Level Dynamics*

There are three predominant sea-level regimes for the middle to late Quaternary Sunda Shelf (Fig. 2) (Morley 2012). The most commonly recurring scenario ( $\pm 55\%$  of last million



years, Fig. 2b) is of periods with sea levels 40-50m below current levels, around half the current shelf emergent, and evergreen rainforests extending from Borneo to Sumatra. The second most common scenario ( $\pm 37\%$  of last million years, Fig. 2a) is of periods with very low sea levels such as the LGM. Vegetation characteristic of seasonal climates was very widespread, though may not have formed a continuous north-to-south corridor for every glacial maximum, and the exposed sandy soils of the shelf may have acted as a substantial barrier to dispersal (Slik et al. 2011). Seasonality of climate may have varied between glacial maxima, suggested by the occurrence of certain mammalian fossils in Java. Faunas from the penultimate and older glacials included many large mammals requiring open woodland (van den Bergh et al. 2001), whereas there were no such immigrants during the LGM for non-forest species. This suggests that opportunities for migration across the Sunda Shelf may have been inhibited during the LGM, while more open vegetation types perhaps with a corridor of semi-evergreen forests but without true savanna, may have been present during earlier glacials (Cannon et al. 2009). Examination of floristic elements in Java suggests that there are semi-evergreen elements that are common to Java and Indochina, but there are no clear Indochinese deciduous forest elements in Java. This suggests that there may have been a dispersal pathway for semi-evergreen elements during the LGM but not for elements requiring more seasonality of climate. From the faunal perspective, the mammals from the Javanese last glacial do not include big non-forest browsers except elephants. Again, this suggests limited dispersal opportunities for open-vegetation browsers. However, the presence of large open-vegetation browsers from earlier glacial intervals in Java suggests that more opportunities for browsers may have occurred earlier in the Quaternary, but due to the floristic differences with Indochina it is thought

unlikely that there was a continuous ‘savanna corridor’. The third climatic scenario is represented by the present day (Fig. 2c) with high sea levels and evergreen rainforests extending from the Kra Isthmus to West Java, including Borneo, but importantly occurring for just 8% of the last million years, emphasizing the ‘refugial’ nature of present day SE Asian rainforests and fauna (Cannon et al. 2009).

These glacial events are thought to have facilitated repeated biotic range expansions between Sumatra, the Thai-Malay Peninsula, Borneo and Java during low sea levels, followed by vicariance as sea levels subsequently rose (Cannon et al. 2009; Gower et al. 2012; de Bruyn et al. 2013). While fossil data could contribute to our understanding of these dynamics, the terrestrial paleontological record from Quaternary SE Asia is still very poorly understood. Though Quaternary sites are relatively abundant throughout the region (Louys and Meijaard 2010), these are mostly represented by cave sites, with often poorly constrained ages. In addition, many taxa have not yet been studied by specialists and much controversy remains regarding taxonomic delineation. Borneo and Sumatra have yielded few terrestrial fossil remains apart from the famous late Pleistocene Niah and Madai Cave sites. There have been a few isolated Proboscidean finds from Samarinda, Brunei and Sarawak, but the exact stratigraphic origin is invariably poorly known (Hooijer 1952; Cranbrook 2010). On the other side of ‘Wallace’s Line’ in the realm of Wallacea, increasing efforts in fossil vertebrate studies were carried out during the second half of the 20th century culminating in the discovery of an endemic late Pleistocene hominin, *Homo floresiensis*, on the oceanic island of Flores (Brown et al. 2004).

The majority of datasets examined here show ‘Out-of-Borneo’ and ‘Out-of-Indochina’ signatures for the Plio-Pleistocene period, while colonization routes from both

Sumatra and the Thai-Malay Peninsula were substantially fewer (Fig. 4, Table 3). Borneo and Indochina have thus been *the* major refugia for Sundaland over at least the last 5 Ma, and likely earlier (see Neogene discussion above).

## DISCUSSION

Given the taxonomic coverage of SE Asia, our phylogenetic meta-analyses of diverse faunal and floral groups identified initiation of diversification during the Jurassic in plants, and insects & spiders, while all other groups examined diversified later during the Cenozoic (Fig. S5-S13, Table S1), consistent with the fossil record. Fossils indicate that a diverse mammalian fauna existed during the Eocene and Oligocene, including various primates, flying lemurs, mustelids, ruminants, and rodents in Myanmar and Thailand (e.g., Marivaux et al. 2004, 2006; Peigné et al. 2006; Beard et al. 2009). There are, however, few Paleogene mammal fossils from island SE Asia, and some of these are of uncertain provenance (e.g., an Eocene anthracothere found on Timor Island and a putative member of the artiodactyl family Haplobunodontidae found in western Borneo; see Ducrocq 1996).

Interestingly, a few datasets show pre-Oligocene diversification events from the Philippines (excluding Palawan) and Java. However, given the small number of taxa with such patterns, dating issues, extinction and range evolution seem more likely at present than a major conflict with the geological evidence. Colonization routes and ancestral areas inferred from meta-analyses of phylogenetic data show that Indochina and Borneo were already major evolutionary hotspots and sources of diversification in the pre-Miocene (Fig. 4, Fig. S14 and S15, Table 2 and 4). These results most likely reflect the paleogeographic history of Borneo and Indochina. Borneo is the largest landmass of the former Sundaland,

and was less affected by sea-level changes compared to other islands like Sumatra and Java, which were largely submerged during the Miocene or even later (Fig. 1). The presence of extended emergent areas including extensive rainforest (Fig. 2) through all time-periods examined here may explain the elevated levels of *in situ* diversification and emigration observed for Indochina and Borneo (Fig. 5, see below).

#### *Caveat on Phylogenetic Meta-Analyses*

Only a few colonizations from the islands within ‘Wallacea excl. Sulawesi’, Palawan and the area east of Wallacea (i.e., mainly New Guinea and Australia) were inferred conducting ancestral area estimation (Fig. 4, Fig. S14 and S15), consistent with a recently published meta-analysis focusing on Sulawesi and Wallacea, which identified that the majority of colonization events were derived from Sundaland, not within or east of Wallacea (Stelbrink et al. 2012). It seems very unlikely that comparatively small and young islands such as the Lesser Sunda Islands and Palawan might have played a major role as refugia and/or colonization sources throughout the timeframe under investigation here, when comparing them with larger, older landmasses such as Borneo. However, one might assume that the lack of particular dispersal routes is a result of weak taxonomic coverage, because some islands are more difficult to access and/or research permits are difficult to procure and therefore taxa from these areas are missing in several phylogenetic studies. One area that is likely to be under-represented in phylogenetic studies is New Guinea, and current evidence suggests this landmass has played a significant role in speciation for the wider region (e.g., Balke 2009; Jönsson et al. 2011; Müller et al. 2013). Similarly, two recent herpetological studies suggest that Palawan may have been more important in the biogeographic history of

SE Asia than generally recognized, acting as a raft for mainland SE Asian lineages when separating from the mainland c. 30 Ma, and thus facilitated dispersal across the Philippines, Borneo, Sulawesi and the Thai-Malay Peninsula during the Miocene (so-called ‘Palawan Ark Hypothesis’; e.g. Blackburn et al. 2010; Siler et al. 2012). However, for most Philippine species groups, Palawan apparently did not function as a dispersal gateway between Borneo and the rest of the Philippines (see e.g., Brown et al. 2013; cf. also Fig. 4). Some species likely dispersed from Borneo to the Philippines (except Palawan) through the Sulu Archipelago (Oliveros and Moyle 2010), while others dispersed from Indochina, probably via Taiwan (Esselstyn and Oliveros 2010). Finally, 15 of 61 datasets analyzed here are based on intraspecific data, while the remainder focus on clades containing multiple species (Table 1). This could result in relatively young divergence times for these intraspecific datasets compared with those datasets studying taxa of a higher taxonomic level (interspecific, intergeneric).

#### *Caveat on Plant and Mammal Distributional Modelling*

SE Asia is a global hotspot for mammal species richness (Catullo et al. 2008; Rondinini et al. 2011). Roughly one quarter of global mammal taxa occur in this area, with many new families and species, which have only been discovered recently (Jenkins et al. 2005; Musser et al. 2005). The region is also a global center of plant richness, especially in northern Borneo and New Guinea. Overall, our study area includes 4 biodiversity hotspots (Sundaland, Wallacea, Philippines, Indo-Burma) as defined by Myers et al. (2000), more than 56,000 plant species are present in the area, of which 52% are endemic to the 4 biodiversity hotspots mentioned (Myers et al. 2000).

We caution against an over-interpretation of the species richness map (Fig. 6), as both are probably overestimating actual species distribution. In fact, increasingly, lowland forest areas in SE Asia are characterized by the so-called ‘empty forest syndrome’, with large animals (mainly primates, carnivores and ungulates) that are extinct in vast areas of their former distribution range because of commercial hunting, even if suitable habitat is still present (Redford 1992; Milner-Gulland et al. 2003; Corlett 2007). Similar issues characterize also the model of plant species richness (Fig. 6). SE Asia is characterized by the highest rate of deforestation globally (Achard et al. 2002; Wilcove et al. 2013), and primary vegetation losses equate to 92.2% of the original extent for Sundaland, 85% for Wallacea, 97% for the Philippines, and 95.1% for Indo-Burma (Myers et al. 2000) in recent decades. If all forest types are included, some 50% of Borneo remains forested (Miettinen et al. 2011) but annual losses continue (1.3%/year, Miettinen et al. 2011; 3%/year for peat swamp forests only, Wilcove et al. 2013). Given that the plant species richness map has been calculated from a database using statistical models, data points, and bioclimatic variables without considering recent annual deforestation and other variables that are likely important determinants of species distributions (e.g., species interactions, microclimatic conditions, etc.), the output map may overestimate the current vascular plants species richness. Importantly though, if absolute values were too high, the proportion and thus species diversity rankings for each area should remain consistent.

Moreover, the two maps of species richness (Fig. 6) actually provide different information. While the model of plant species richness gives a ‘biogeographic’ estimate of the potential species richness, mammal species richness is more similar to the actual species richness due to issues related to the historical alteration of the species distribution ranges

(e.g., habitat destruction, hunting, and other human related activities may have altered the original species distribution). Thus, species richness for mammals may also reflect historical/human-related factors, as well as biogeography.

#### *Indochina and Borneo: Major Evolutionary Hotspots*

Geological and climatic processes associated with the early Miocene collision (see above) may have effectively divided the biota into two major Miocene areas north and south of the Thai-Malay Peninsula: namely, Indochina and (an extended) Borneo (Fig. 1, 15 Ma).

Consistent with an expectation of *in situ* diversification in refugia and thus the accumulation of lineages during the global middle Miocene Climate and sea-level maximum, our analyses of phylogenetic diversification events show an increase in the rate of diversification consistent with this time period (Fig. 5 and Fig. S13, Table S4). This is most evident for Indochina and Borneo, suggesting a role for global Miocene sea-level fluctuations and associated refugia development on biotic evolution.

SE Asia's biodiversity has evidently been shaped by both geological and climatic drivers resulting in an extremely complicated history of past land (re)connections, including micro-terrane movements, with a young age of many diversification events evident in the phylogenetic meta-analyses, a high proportion of which occur in the latest Pliocene and Pleistocene (Fig. S13; see also Stelbrink et al. 2012 for details on Sulawesi). During the Pliocene, Borneo essentially maintained its earlier configuration. The Barisan Mountains of Sumatra continued to develop to their present altitudes (Fig. 1, 5 Ma); the Sumatran coastal plain expanded and at times of low sea level may have been contiguous with the Thai-Malay Peninsula. The island of Java essentially formed during this period and floristic

differences from Sumatra as emphasized by van Welzen et al. (2011) probably relate to its persistently drier climate and young geological age. A substantial increase in the colonization of other Greater Sunda Islands and the Philippines from Borneo and Indochina in the Miocene and even more pronounced in the Plio-Pleistocene (Fig. 4 and Table 3), although representing a much shorter timeframe compared to the Miocene. An increase in colonization events between adjacent areas of Sundaland (see Fig. 4 and Table 3), is potentially contemporaneous with the successive (re)emergence of land in these areas and cyclical climate-induced connectivity (Fig. 2). Consistent with these young emergent ages, ancestral area estimation generally show less frequent colonization events out of Java and Sumatra (compared to Borneo and Indochina), but higher levels of immigration events into these areas (Fig. 4 and Table 3). Hence, climate-forced sea-level changes facilitating inter-area connection-disconnection events have played a substantial role in providing new dispersal ‘corridors’ for such lineage dispersal across SE Asia.

Indochina, and Borneo especially, the two largest SE Asian areas examined here with the longest emergent histories, have been major diversification hotspots through time and key sources for lineage dispersal across the region (Fig. 4, Table 3). Such larger areas may also have experienced comparatively fewer extinction events due to presence of additional putative refugia. Our analyses suggest length of emergent history has been a more important factor than current area size for both total number of lineages and levels of emigration events for Indochina and Borneo (Table S4), however, future analytical approaches should consider changes in area size since emergence. For example, analytical advances to test for both ancestral area estimation and geography-diversification correlations have recently become available (e.g., R package BioGeoBEARS, Matzke



2013; GeoSSE, Goldberg et al. 2011). Moreover, our analyses show that the prolonged emergent history of Indochina and Borneo initially facilitated the accumulation of lineages in both biogeographic areas (Fig. 5 and Fig. S13, Table S1 and S5), resulting in increased levels of emigration across SE Asia when remaining islands (re)emerged over the course of Mio- and Plio-Pleistocene sea-level fluctuations (Fig. 2 and 4, Table 3 and Table S3).

In summary, we find evidence for all four hypotheses set out in the Introduction, namely: levels of *in situ* diversification events reflected emergent ages (but not current sizes) of our pre-defined areas; areas with recent histories of sea-level inundation (e.g., Sumatra and Java) were characterized by higher levels of (recent) immigration than emigration; number of colonization events were greater between adjacent areas compared to more distant areas; and finally, an increase in diversification rates was evident during the Plio-Pleistocene, likely resulting from increased connection-disconnection events related to sea-level fluctuations.

While the importance of Indochina to the evolutionary history of SE Asian biodiversity has long been recognized (Myers et al. 2000; Lohman et al. 2011 and references therein), Borneo, embedded in one of the world's most exceptionally diverse biodiversity hotspots (Sundaland; see Myers et al. 2000) is clearly also of extraordinary importance to the SE Asian biota, as shown here across several taxonomic groups through all time periods investigated. Much of Borneo's remarkable diversity, including remnant populations of conservation icons such as Sumatran rhinoceros (*Dicerorhinus sumatrensis*) and Bornean orangutan (*Pongo pygmaeus*), is under extreme and sustained pressure from anthropogenic habitat conversion (Miettinen et al. 2011) and unsustainable use (Koh and Sodhi 2010; Wilcove et al. 2013).

These results conclusively drive home the importance of maintaining extensive high priority conservation areas in Borneo as envisaged under the current ‘Heart of Borneo’ agreement between Indonesia, Malaysia and Brunei. However, the agreement fails to adequately protect extensive areas of lowland rainforest, which harbor the highest levels of diversity. Losing further large areas of forested land to development in this region will result in the irreplaceable loss of the primary refuge area for the entire Sunda Shelf region.

#### SUPPLEMENTARY MATERIAL

Supplementary Material (alignment fasta files, BEAST xml files, Lagrange input and output files) is available online in the Dryad data repository at <http://datadryad.org>, doi: 10.5061/dryad.67s40 including Supplementary Information/Text, Supplementary Tables S1-S4, Supplementary Figures S1-S15.

#### FUNDING

I.M. acknowledges funding from the Australian Research Council. G.v.d.B. was funded by Australian Research Council Future Fellowship No. FT100100384, and thanks the Geological Survey Institute in Bandung, and the Archaeological Research and Development Centre in Jakarta (ARKENAS) for longstanding collaborations in the field. T.v.R. was funded by German Research Council (DFG) grant RI 1738/4-1.

#### ACKNOWLEDGEMENTS

We are indebted to Ladislav Bocak (Czech Republic), Sebastian Klaus (Germany), David Lohman (USA), Marie-Lilith Patou (France), Hsi-Te Shih (Taiwan) and T.N.C. Vidya (India) for providing data for the ancestral estimation analyses. We thank Adrian Patterson, Jim A. McGuire, Nick Matzke, K.M. Wong and several additional anonymous reviewers for comments and suggestions that greatly improved the manuscript. M.d.B., B.S., T.v.R. conceived and designed the study; all authors collected and analyzed the data, and contributed to writing the paper.

## REFERENCES

- Achard F., Eva H.D., Stibig H.-J., Mayaux P., Gallego J., Richards T., Malingreau J.-P. 2002. Determination of deforestation rates of the world's humid tropical forests. *Science* 297:999–1002.
- Aldrich J.B., Rinehart G.P., Ridwan S., Schuepbach M.A. 1996. Paleogene basin architecture of the Sunda and Asri Basins and associated non-marine sequence stratigraphy. *Proc. Int. Symp. Seq. Stratigr. S.E. Asia*. 261–287.
- Armstrong R.A. (2014) When to use the Bonferroni correction. *Ophthalmic & Physiological Optics*, In Press. doi: 10.1111/opo.12131
- Balke M. 2009. New Guinea highland origin of a widespread arthropod supertramp. *Proc. R. Soc. London B* 276:2359–2367.
- Beard K.C., Marivaux L., Chaimanee Y., Jaeger J.-J., Marandat B., Tafforeau P., Soe A.N., Tun S.T., Kyaw A.A. 2009. A new primate from the Eocene Pondaung Formation of Myanmar and the monophyly of Burmese amphipithecids. *Proc. R. Soc. London B* 276:3285–3294.
- Blackburn D.C., Bickford D.P., Diesmos A.C., Iskandar D.T., Brown R.M. 2010. An ancient origin for the enigmatic flat-headed frogs (Bombinatoridae: *Barbourula*) from the islands of Southeast Asia. *PLoS ONE* 5:e12090.
- Brown P., Sutikna T., Morwood M.J., Soejono R.P., Jatmiko, Saptomo E.W., Due R.A. 2004. A new small-bodied hominin from the Late Pleistocene of Flores, Indonesia. *Nature* 431:1055–1061.
- Brown R.M., Siler C.D., Oliveros C.H., Esselstyn J. a., Diesmos A.C., Hosner P. a., Linkem C.W., Barley A.J., Oaks J.R., Sanguila M.B., Welton L.J., Blackburn D.C., Moyle R.G., Townsend Peterson A., Alcalá A.C. 2013. Evolutionary processes of diversification in a model island archipelago. *Annu. Rev. Ecol. Evol. Syst.* 44:411–435.
- Cannon C.H., Morley R.J., Bush A.B.G. 2009. The current refugial rainforests of Sundaland are unrepresentative of their biogeographic past and highly vulnerable to disturbance. *Proc. Natl. Acad. Sci. USA* 106:11188–11193.
- Catullo G., Masi M., Falcucci A., Maiorano L., Rondinini C., Boitani L. 2008. A gap analysis of Southeast Asian mammals based on habitat suitability models. *Biol. Conserv.* 141:2730–2744.

- Corlett R.T. 2007. The impact of hunting on the mammalian fauna of tropical Asian forests. *Biotropica* 39:292–303.
- Earl of Cranbrook 2010. Late quaternary turnover of mammals in Borneo: the zooarchaeological record. *Biodivers. Conserv.* 19:373–391.
- de Bruyn M., Rüber L., Nylinder S., Stelbrink B., Lovejoy N.R., Lavoué S., Tan H.H., Nugroho E., Wowor D., Ng P.K.L., Siti Azizah M.N., von Rintelen T., Hall R., Carvalho G.R. 2013. Paleo-drainage basin connectivity predicts evolutionary relationships across three Southeast Asian biodiversity hotspots. *Syst. Biol.* 62:398–410.
- Drummond A.J., Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* 7:214.
- Ducrocq S. 1996. The Eocene terrestrial mammal from Timor, Indonesia. *Geol. Mag.* 133:763–766.
- Esselstyn J.A., Oliveros C.H. 2010. Colonization of the Philippines from Taiwan: a multi-locus test of the biogeographic and phylogenetic relationships of isolated populations of shrews. *J. Biogeogr.* 37:1504–1514.
- Goldberg E.E., Lancaster L.T., Ree R.H. 2011. Phylogenetic inference of reciprocal effects between geographic range evolution and diversification. *Syst. Biol.* 60:451–465.
- Gower D., Johnson K.G., Rosen B.R., Richardson J., Rüber L., Williams S.T. 2012. Biotic evolution and environmental change in Southeast Asia. Cambridge: Cambridge University Press.
- Grummer J.A., Bryson Jr. R.W., Reeder T.W. 2014. Species delimitation using Bayes factors: simulations and application to the *Sceloporus scalaris* species group (Squamata: Phrynosomatidae). *Syst. Biol.* 63:119–133.
- Hall R. 2009. Southeast Asia's changing palaeogeography. *Blumea* 54:148–161.
- Hall R. 2011. Australia-SE Asia collision: plate tectonics and crustal flow. In: Hall R., Cottam M.A., Wilson M.E.J., editors. *The Southeast Asian Gateway: history and tectonics of Australia-Asia collision*. London: The Geological Society of London. p. 75–109.
- Hall R. 2012a. Late Jurassic-Cenozoic reconstructions of the Indonesian region and the Indian Ocean. *Tectonophysics.* 570-571:1–41.

- Hall R. 2012b. Sundaland and Wallacea: geology, plate tectonics and palaeogeography. In: Gower D.J., Johnson K.G., 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.
- Hall R. 2013. The palaeogeography of Sundaland and Wallacea since the Late Jurassic. *J. Limnol.* 72:1–17.
- Hanebuth T.J.J., Voris H.K., Yokoyama Y., Saito Y., Okuno J. 2011. Formation and fate of sedimentary depocentres on Southeast Asia's Sunda Shelf over the past sea-level cycle and biogeographic implications. *Earth-Science Rev.* 104:92–110.
- Harley M.M., Morley R.J. 1995. Ultrastructural studies of some fossil and extant palm pollen, and the reconstruction of the biogeographical history of subtribes Iguanurinae and Calaminae. *Rev. Palaeobot. Palynol.* 85:153–182.
- Hooijer D.A. 1952. *Palaeoloxodon cf. namadicus* (Falconer et Cautley) from Borneo. *Proc. K. Ned. Akad. van Wet. Ser. B.* 55:395–398.
- Jenkins P.D., Kilpatrick C.W., Robinson M.F., Timmins R.J. 2005. Morphological and molecular investigations of a new family, genus and species of rodent (Mammalia: Rodentia: Hystricognatha) from Lao PDR. *Syst. Biodivers.* 2:419–454.
- Jønsson K.A., Fabre P.H., Ricklefs R.E., Fjeldså J. 2011. Major global radiation of corvid birds originated in the proto-Papuan archipelago. *Proc. Natl. Acad. Sci. USA* 108:2328–2333.
- Katoh K., Toh H. 2008. Recent developments in the MAFFT multiple sequence alignment program. *Brief. Bioinform.* 9:286–298.
- Klaus S., Schubart C.D., Streit B., Pfenninger M. 2010. When Indian crabs were not yet Asian - biogeographic evidence for Eocene proximity of India and Southeast Asia. *BMC Evol. Biol.* 10:287.
- Koh L.P., Sodhi N.S. 2010. Conserving Southeast Asia's imperiled biodiversity: scientific, management, and policy challenges. *Biodivers. Conserv.* 19:913–917.
- Kreft H., Jetz W. 2007. Global patterns and determinants of vascular plant diversity. *Proc. Natl. Acad. Sci. USA* 104:5925–5930.
- Li J.-T., Li Y., Klaus S., Rao D.-Q., Hillis D.M., Zhang Y.-P. 2013. Diversification of rhacophorid frogs provides evidence for accelerated faunal exchange between India and Eurasia during the Oligocene. *Proc. Natl. Acad. Sci. USA* 110:3441–3446.

- Lohman D.J., de Bruyn M., Page T., von Rintelen K., Hall R., Ng P.K.L., Shih H.T., Carvalho G.R., von Rintelen T. 2011. Biogeography of the Indo-Australian archipelago. *Annu. Rev. Ecol. Evol. Syst.* 42:205–226.
- Louys J., Meijaard E. 2010. Palaeoecology of Southeast Asian megafauna-bearing sites from the Pleistocene and a review of environmental changes in the region. *J. Biogeogr.* 37:1432–1449.
- Marivaux L., Bocat L., Chaimanee Y., Jaeger J.-J., Marandat B., Srisuk P., Tafforeau P., Yamee C., Welcomme J.-L. 2006. Cynocephalid dermopterans from the Palaeogene of South Asia (Thailand, Myanmar and Pakistan): systematic, evolutionary and palaeobiogeographic implications. *Zool. Scr.* 35:395–420.
- Marivaux L., Chaimanee Y., Yamee C., Srisuk P., Jaeger J.-J. 2004. Discovery of *Fallomus ladakhensis* Nanda & Sahni, 1998 (Rodentia, Diatomyidae) in the lignites of Nong Ya Plong (Phetchaburi Province, Thailand): systematic, biochronological and paleoenvironmental implications. *Geodiversitas* 26:493–507.
- Matzke N.J. 2013. Probabilistic historical biogeography: new models for founder-event speciation, imperfect detection, and fossils allow improved accuracy and model-testing. *Front. Biogeogr.* 5:242–248.
- Mayr E. 1944. Wallace's line in the light of recent zoogeographic studies. *Q. Rev. Biol.* 19:1–14.
- Metcalf I. 2011a. Tectonic framework and Phanerozoic evolution of Sundaland. *Gondwana Res.* 19:3–21.
- Metcalf I. 2011b. Palaeozoic-Mesozoic history of SE Asia. In: Hall R., Cottam M.A., Wilson M.E.J., editors. *The Southeast Asian Gateway: history and tectonics of Australia-Asia collision*. London: The Geological Society of London. p. 7–35.
- Miettinen J., Shi C., Liew S.C. 2011. Deforestation rates in insular Southeast Asia between 2000 and 2010. *Glob. Chang. Biol.* 17:2261–2270.
- Milner-Gulland E.J., Bennett E.L., SCB-2002-Annual-Meeting-Wild-Meat-Group. 2003. Wild meat: the bigger picture. *Trends Ecol. Evol.* 18:351–357.
- Moulds P.J. 1989. Development of the Bengkalis Depression, Central Sumatra and its subsequent deformation – a model for other Sumatran grabens? 18th Annu. Conv. Proc. (Indones. Pet. Assoc.). 217–245.
- Moran M.D. 2003. Arguments for rejecting the sequential Bonferroni in ecological studies. *Oikos* 100:403–405.

- Morley R.J., Morley H.P. 2011. Neogene climate history of the Makassar Straits. The Southeast Asian Gateway: history and tectonics of Australia-Asia collision. London: Geological Society of London, Special Publication. p. 319–332.
- Morley R.J., Morley H.P. 2013. Mid Cenozoic freshwater wetlands of the Sunda region. *J. Limnol.* 72:18–35.
- Morley R.J., Swiecicki T., Pham D.T.T. 2012. A sequence stratigraphic framework for the Sunda region, based on integration of biostratigraphic, lithological and seismic data from Nam Con Son Basin, Vietnam. *Proc. Indones. Pet. Assoc. 35th Annual:IPA11–G–002.*
- Morley R.J. 1996. Biostratigraphic characterization of systems tracts in Tertiary sedimentary basins. *Proc. Int. Symp. Seq. Stratigr. S.E. Asia.* 49–71.
- Morley R.J. 2000. *Origin and evolution of tropical rain forests.* Chichester: John Wiley & Sons.
- Morley R.J. 2006. Cretaceous and Tertiary climate change and the past distribution of megathermal rain forests. In: Bush M., J.R. F., editors. *Tropical rainforest responses to climatic change.* Springer Praxis. p. 1–54.
- Morley R.J. 2012. A review of the Cenozoic palaeoclimate history of Southeast Asia. In: Gower D., Johnson K.G., Rosen B.R., Richardson J., Rüber L., Williams S.T., editors. *Biotic evolution and environmental change in Southeast Asia.* Cambridge: Cambridge University Press. p. 79–114.
- Müller C.J., Matos-Maraví P.F., Beheregaray L.B. 2013. Delving into *Delias* Hübner (Lepidoptera: Pieridae): fine-scale biogeography, phylogenetics and systematics of the world's largest butterfly genus. *J. Biogeogr.* 40:881–893.
- Musser G.G., Smith A.L., Robinson M.F., Lunde D.P. 2005. Description of a new genus and species of rodent (Murinae, Muridae, Rodentia) from the Khammouan Limestone National Biodiversity Conservation Area in Lao PDR. *Am. Museum Novit.* 3497:1–32.
- Myers N., Mittermeier R.A., Mittermeier C.G., da Fonseca G.A.B., Kent J. 2000. Biodiversity hotspots for conservation priorities. *Nature* 403:853–858.
- Oliveros C.H., Moyle R.G. 2010. Origin and diversification of Philippine bulbuls. *Mol. Phylogenet. Evol.* 54:822–832.



- Peigné S., Chaimanee Y., Yamee C., Srisuk P., Marandat B., Jaeger J.-J. 2006. A new member of the Mustelida (Mammalia: Carnivora) from the Paleogene of Southern Asia. *J. Vertebr. Paleontol.* 26:788–793.
- Perneger T. V. 1998. What's wrong with Bonferroni adjustments. *BMJ* 316:1236–1238.
- Posada D. 2008. jModelTest: Phylogenetic model averaging. *Mol. Biol. Evol.* 25:1253–1256.
- Rambaut A., Drummond A.J. 2007. Tracer v. 1.5, Available from <http://beast.bio.ed.ac.uk/Tracer>
- Redford K.H. 1992. The empty forest. *Bioscience* 42:412–422.
- 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.
- Richardson J.E., Costion C.M., Muellner A.N. 2012. The Malesian floristic interchange: plant migration patterns across Wallace's Line. In: Gower D., Johnson K.G., Rosen B.R., Richardson J., Rüber L., Williams S.T., editors. *Biotic evolution and environmental change in Southeast Asia*. Cambridge: Cambridge University Press. p. 138–163.
- Rondinini C., Di Marco M., Chiozza F., Santulli G., Baisero D., Visconti P., Hoffmann M., Schipper J., Stuart S.N., Tognelli M.F., Amori G., Falcucci A., Maiorano L., Boitani L. 2011. Global habitat suitability models of terrestrial mammals. *Philos. Trans. R. Soc. London B* 366:2633–2641.
- Shih H.T., Yeo D.C.J., Ng P.K.L. 2009. The collision of the Indian plate with Asia: molecular evidence for its impact on the phylogeny of freshwater crabs (Brachyura: Potamidae). *J. Biogeogr.* 36:703–719.
- Shoup R.C., Morley R.J., Swiecicki T., Clark S.J. 2012. Tectono-stratigraphic framework and Tertiary paleogeography of Southeast Asia: Gulf of Thailand to South Vietnam Shelf. *AAPG Search Discov. Artic.* #30246 (posted Sept 24 2012).
- Siler C.D., Oaks J.R., Welton L.J., Linkem C.W., Swab J.C., Diesmos A.C., Brown R.M. 2012. Did geckos ride the Palawan raft to the Philippines? *J. Biogeogr.* 39:1217–1234.

- Simpson G.G. 1977. Too many lines; the limits of the Oriental and Australian zoogeographic regions. *Proc. Am. Philos. Soc.* 121:107–120.
- Slik J.W.F., Aiba S.-I., Bastian M., Brearley F.Q., Cannon C.H., Eichhorn K.A.O., Fredriksson G., Kartawinata K., Laumonier Y., Mansor A., Marjokorpi A., Meijaard E., Morley R.J., Nagamasu H., Nilus R., Nurtjahya E., Payne J., Permana A., Poulsen A.D., Raes N., Riswan S., Van Schaik C.P., Sheil D., Sidiyasa K., Suzuki E., van Valkenburg J.L.C.H., Webb C.O., Wich S., Yoneda T., Zakaria R., Zweifel N. 2011. Soils on exposed Sunda Shelf shaped biogeographic patterns in the equatorial forests of Southeast Asia. *Proc. Natl. Acad. Sci. USA* 108:12343–12347.
- Spakman W., Hall R. 2010. Surface deformation and slab–mantle interaction during Banda arc subduction rollback. *Nat. Geosci.* 3:562–566.
- Stelbrink B., Albrecht C., Hall R., von Rintelen T. 2012. The biogeography of Sulawesi revisited: is there evidence for a vicariant origin of taxa on Wallace’s “anomalous island”? *Evolution* 66:2252–2271.
- Sun X., Wang P. 2005. How old is the Asian monsoon system? – Palaeobotanical records from China. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 222:181–222.
- Turner H., Hovenkamp P., van Welzen P.C. 2001. Biogeography of Southeast Asia and the West Pacific. *J. Biogeogr.* 28:217–230.
- Vakhrameev V.A. 1991. Jurassic and Cretaceous floras and climates of the Earth. Cambridge: Cambridge University Press.
- van den Bergh G.D., de Vos J., Sondaar P.Y. 2001. The late Quaternary palaeogeography of mammal evolution in the Indonesian archipelago. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 171:385–408.
- van Oosterzee P. 1997. Where worlds collide. The Wallace Line. Ithaca: Cornell University Press.
- van Welzen P.C., Parnell J.A.N., Slik J.W.F. 2011. Wallace’s Line and plant distributions: two or three phytogeographical areas and where to group Java? *Biol. J. Linn. Soc.* 103:531–545.
- Voris H.K. 2000. Maps of Pleistocene sea levels in Southeast Asia: shorelines, river systems and time durations. *J. Biogeogr.* 27:1153–1167.
- Wakita K., Metcalfe I. 2005. Ocean plate stratigraphy in East and Southeast Asia. *J. Asian Earth Sci.* 24:679–702.

- Wallace A.R. 1869. *The Malay Archipelago, the land of the orang-utan and the bird of paradise: a narrative of travel with studies of man and nature*. London: Macmillan.
- Whitmore T.C. 1981. *Wallace's line and plate tectonics*. Oxford: Clarendon Press.
- Whitmore T.C. 1987. *Biogeographical evolution of the Malay Archipelago*. Oxford: Clarendon Press.
- Wilcove D.S., Giam X., Edwards D.P., Fisher B., Koh L.P. 2013. Navjot's nightmare revisited: logging, agriculture, and biodiversity in Southeast Asia. *Trends Ecol. Evol.* 28:531–540.
- Witts D., Hall R., Nichols G., Morley R. 2012. A new depositional and provenance model for the Tanjung Formation, Barito Basin, SE Kalimantan, Indonesia. *J. Asian Earth Sci.* 56:77–104.
- Wong K.M. 2011. A biogeographic history of Southeast Asian rainforests. In: Wickneswari R., Cannon C., editors. *Managing the future of Southeast Asia's valuable tropical rainforests*. Dordrecht: Springer Netherlands. p. 21–55.
- Woodruff D.S., Turner L.M. 2009. The Indochinese-Sundaic zoogeographic transition: a description and analysis of terrestrial mammal species distributions. *J. Biogeogr.* 36:803–821.
- Woodruff D.S. 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.
- Xia X., Xie Z. 2001. DAMBE: Data analysis in molecular biology and evolution. *J. Hered.* 92:371–373.
- Zachos J., Pagani M., Sloan L., Thomas E., Billups K. 2001. Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science* 292:393–686.

## Figure Legends

FIGURE 1. Paleogene and Neogene maps of Southeast Asian paleogeography and key habitat availability. Paleogeography of the region from Sundaland to northern Australia and the west Pacific between 30 Ma and 5 Ma overlain on tectonic reconstructions of Hall (2012a) in which the basis for the reconstructions is described in detail (see Materials and Methods for further information).

FIGURE 2. Three Quaternary climate sea-level scenarios for Southeast Asia: a) Scenario 1: 25 ka at lowest sea levels, glacial maximum, Marine Isotope Stage (MIS) 2. b) Scenario 2: 12 ka as sea flooded Sunda Shelf, intermediate sea level, MIS Stages 1, 3, 4, 5a-d. c) Scenario 3: present-day, interglacial high sea level, MIS Stages 1, 5e. Rainforest distribution based on fig. 1a in Cannon et al. (2009). Last Glacial Maximum pollen localities from Morley (2012). d) Land area of tropical evergreen broadleaf forest over last 1 million years, from Cannon et al. (2009); horizontal red lines show times when scenarios 1-3 would likely have occurred (note that the blue line represents the most conservative model but not the most likely one here). Note that Borneo, in particular, has hosted extensive rainforests throughout all periods shown.

FIGURE 3. Exemplary illustration (dataset 33, bush warblers) how colonization routes through time were inferred from both MCC trees (BEAST) and ancestral area estimation (Lagrange; colored circles). The blue time bar (0.16-0.9 Ma) represents the divergence time of the MRCA in Sumatra and is plotted in Fig. S11 and S13 (middle panel). The lineage of interest represents a specimen from area C (Sumatra). This lineage can be traced back to a

node with an ancestral area in A (Indochina) 1.25 Ma (black arrow) and hence a colonization route from area A to area C in the ‘Plio-Pleistocene’ timeframe is visualized in the bottom panel (see Fig. 4 and Fig. S14 and S15).

FIGURE 4. Colonization routes of extant Southeast Asian taxa. a) Colonization routes inferred from ancestral area estimation (DEC model, Lagrange) for lineages among studied taxa in the pre-Miocene, Miocene, and Plio-Pleistocene (see Material and Methods and Fig. 3 for details). Line-color corresponds to inferred area of origin (see inset map), while line thickness corresponds to number of colonizations; multiple colonizations found for a specific source and destination (e.g., Indochina-Sumatra) from a single dataset are counted once only. See Fig. S14 and S15 for group-specific colonization routes. b) Number of emigration (positive bars) and immigration (negative bars) events observed for pre-defined areas in the pre-Miocene, Miocene, and Plio-Pleistocene.

FIGURE 5. Distribution of lineages and *in situ* diversification events among all 61 datasets through time. a) Group-specific (plants, insects & spiders, freshwater crustaceans, freshwater molluscs, freshwater fishes, herpetofauna, birds, and mammals), total combined, and total combined excluding population-based number of lineages across all datasets from each pre-defined area. b) Group-specific (as above), total combined, and total combined excluding population-based number of *in situ* diversification events across all datasets from each pre-defined area.

FIGURE 6. Extant Southeast Asian mammal and plant species richness calculated as the sum of species-specific distribution models. Borneo hosts the highest extant number of mammalian species ( $n=230$ ), and also the highest weighted species richness for both vascular plants and mammals (together with Thai-Malay Peninsula for top 10% of mammals) for SE Asia (see Material and Methods for details). Plant species richness data from Kreft and Jetz (2007), *Copyright (2012) National Academy of Sciences, U.S.A.*

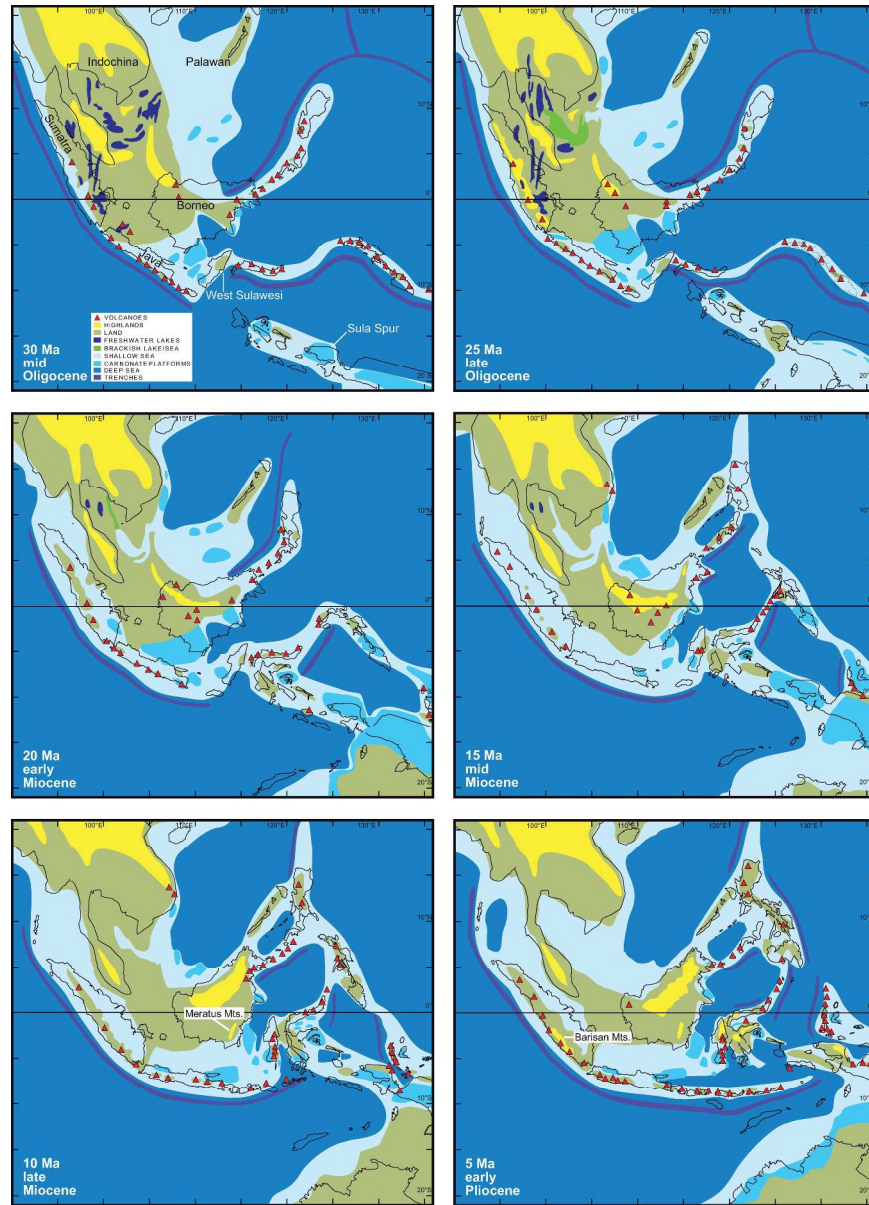


FIGURE 1. Paleogene and Neogene maps of Southeast Asian paleogeography and key habitat availability. Paleogeography of the region from Sundaland to northern Australia and the west Pacific between 30 Ma and 5 Ma overlay on tectonic reconstructions of Hall (2012a) in which the basis for the reconstructions is described in detail (see Materials and Methods for further information).  
 243x337mm (300 x 300 DPI)

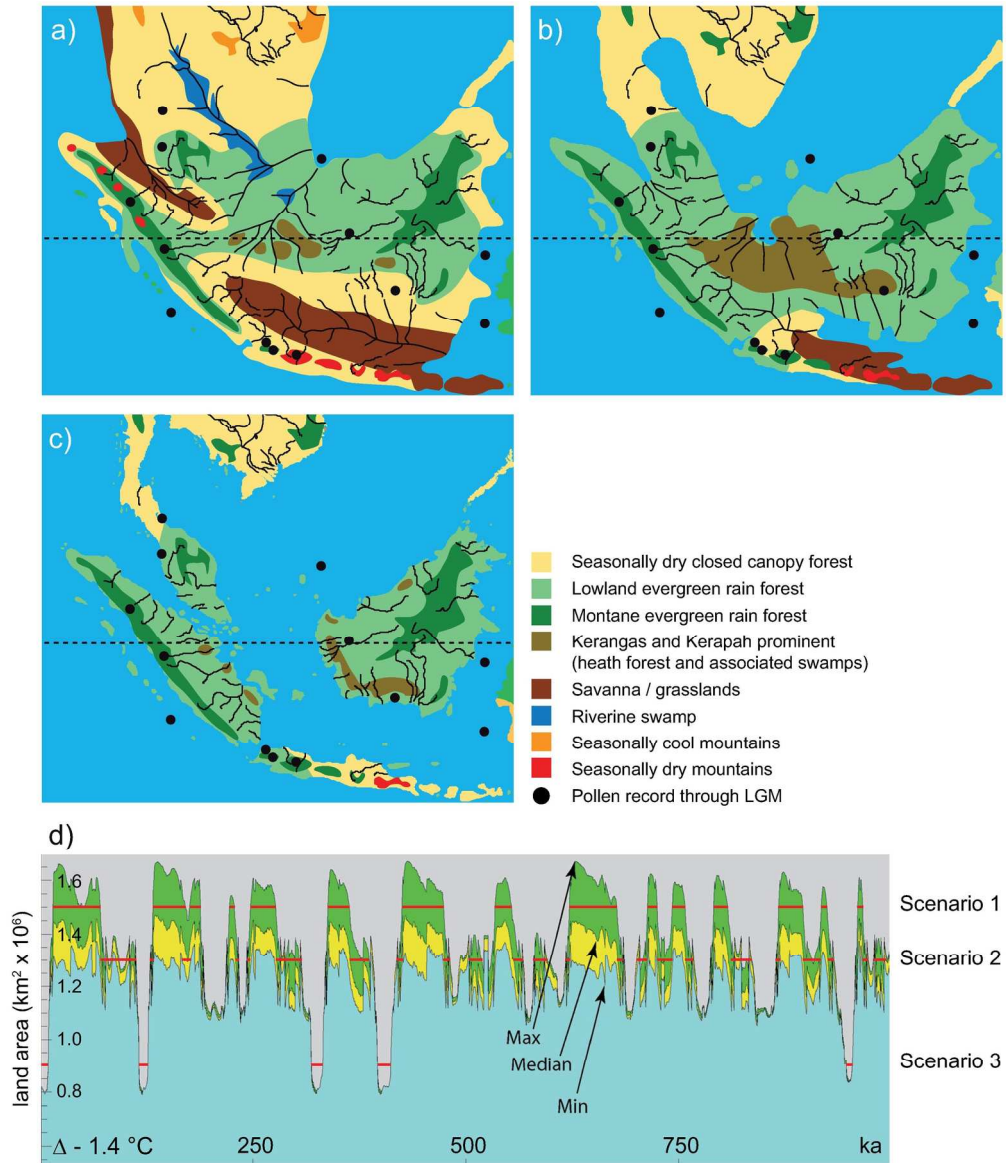


FIGURE 2. Three Quaternary climate sea-level scenarios for Southeast Asia: a) Scenario 1: 25 ka at lowest sea levels, glacial maximum, Marine Isotope Stage (MIS) 2. b) Scenario 2: 12 ka as sea flooded Sunda Shelf, intermediate sea level, MIS Stages 1, 3, 4, 5a-d. c) Scenario 3: present-day, interglacial high sea level, MIS Stages 1, 5e. Rainforest distribution based on fig. 1a in Cannon et al. (2009). Last Glacial Maximum pollen localities from Morley (2012). d) Land area of tropical evergreen broadleaf forest over last 1 million years, from Cannon et al. (2009); horizontal red lines show times when scenarios 1-3 would likely have occurred (note that the blue line represents the most conservative model but not the most likely one here). Note that Borneo, in particular, has hosted extensive rainforests throughout all periods shown.

174x203mm (300 x 300 DPI)



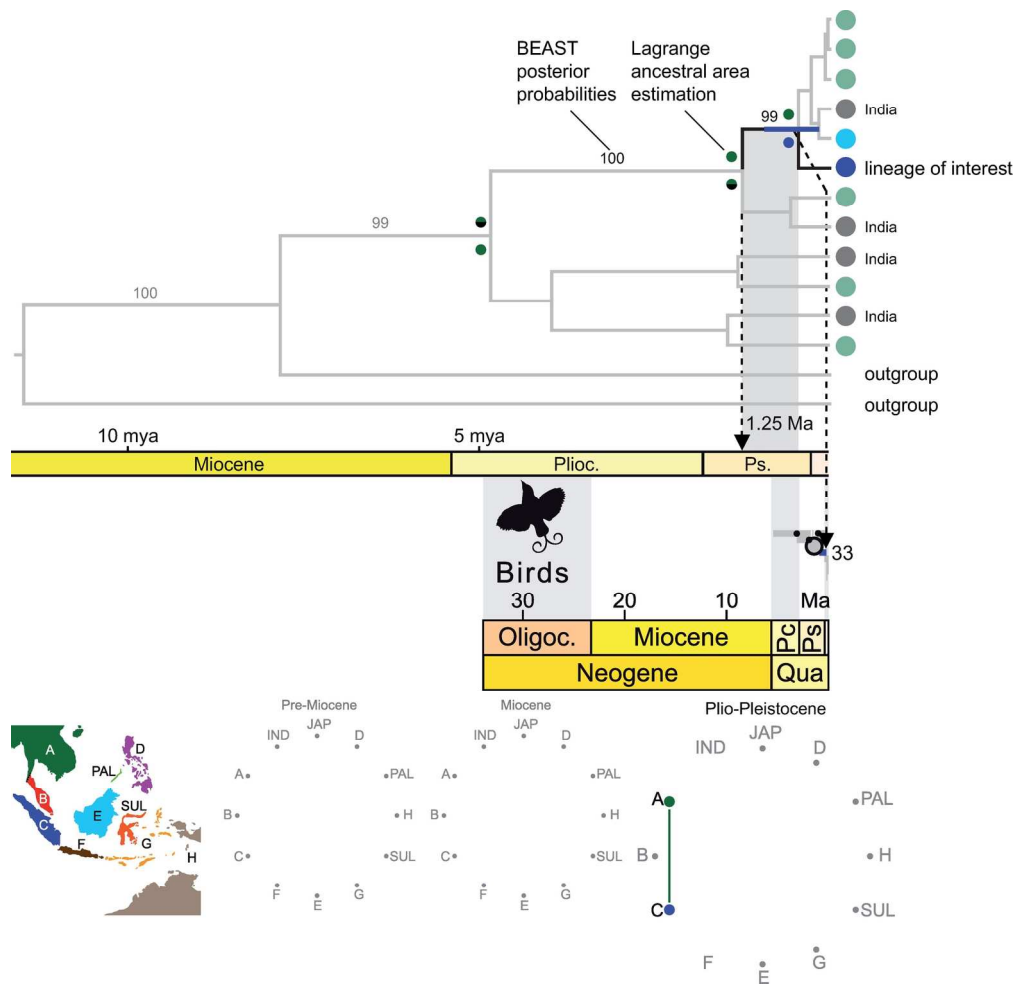


FIGURE 3. Exemplary illustration (dataset 33, bush warblers) how colonization routes through time were inferred from both MCC trees (BEAST) and ancestral area estimation (Lagrange; colored circles). The blue time bar (0.16-0.9 Ma) represents the divergence time of the MRCA in Sumatra and is plotted in Fig. S11 and S13 (middle panel). The lineage of interest represents a specimen from area C (Sumatra). This lineage can be traced back to a node with an ancestral area in A (Indochina) 1.25 Ma (black arrow) and hence a colonization route from area A to area C in the 'Plio-Pleistocene' timeframe is visualized in the bottom panel (see Fig. 4 and Fig. S14 and S15).  
 170x165mm (300 x 300 DPI)

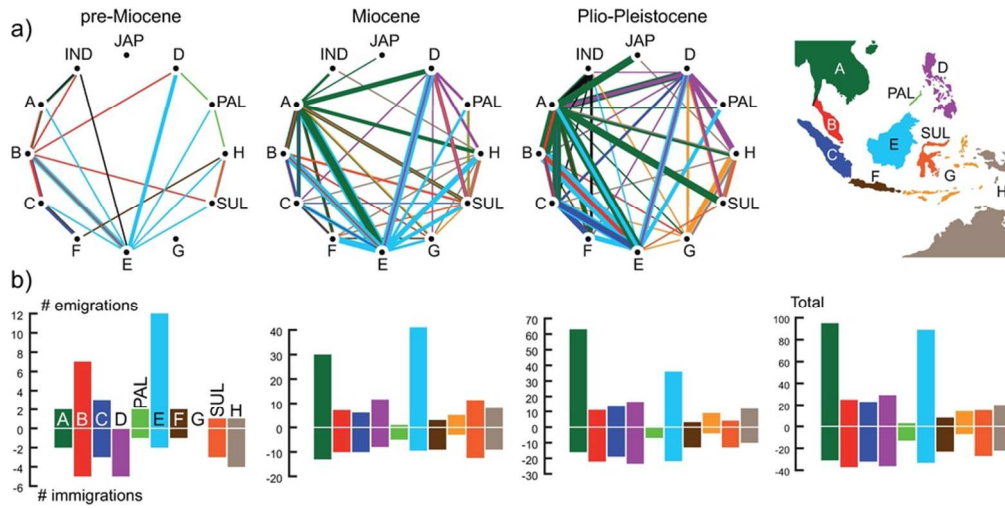


FIGURE 4. Colonization routes of extant Southeast Asian taxa. a) Colonization routes inferred from ancestral area estimation (DEC model, Lagrange) for lineages among studied taxa in the pre-Miocene, Miocene, and Plio-Pleistocene (see Material and Methods and Fig. 3 for details). Line-color corresponds to inferred area of origin (see inset map), while line thickness corresponds to number of colonizations; multiple colonizations found for a specific source and destination (e.g., Indochina-Sumatra) from a single dataset are counted once only. See Fig. S14 and S15 for group-specific colonization routes. b) Number of emigration (positive bars) and immigration (negative bars) events observed for pre-defined areas in the pre-Miocene, Miocene, and Plio-Pleistocene.

87x43mm (300 x 300 DPI)

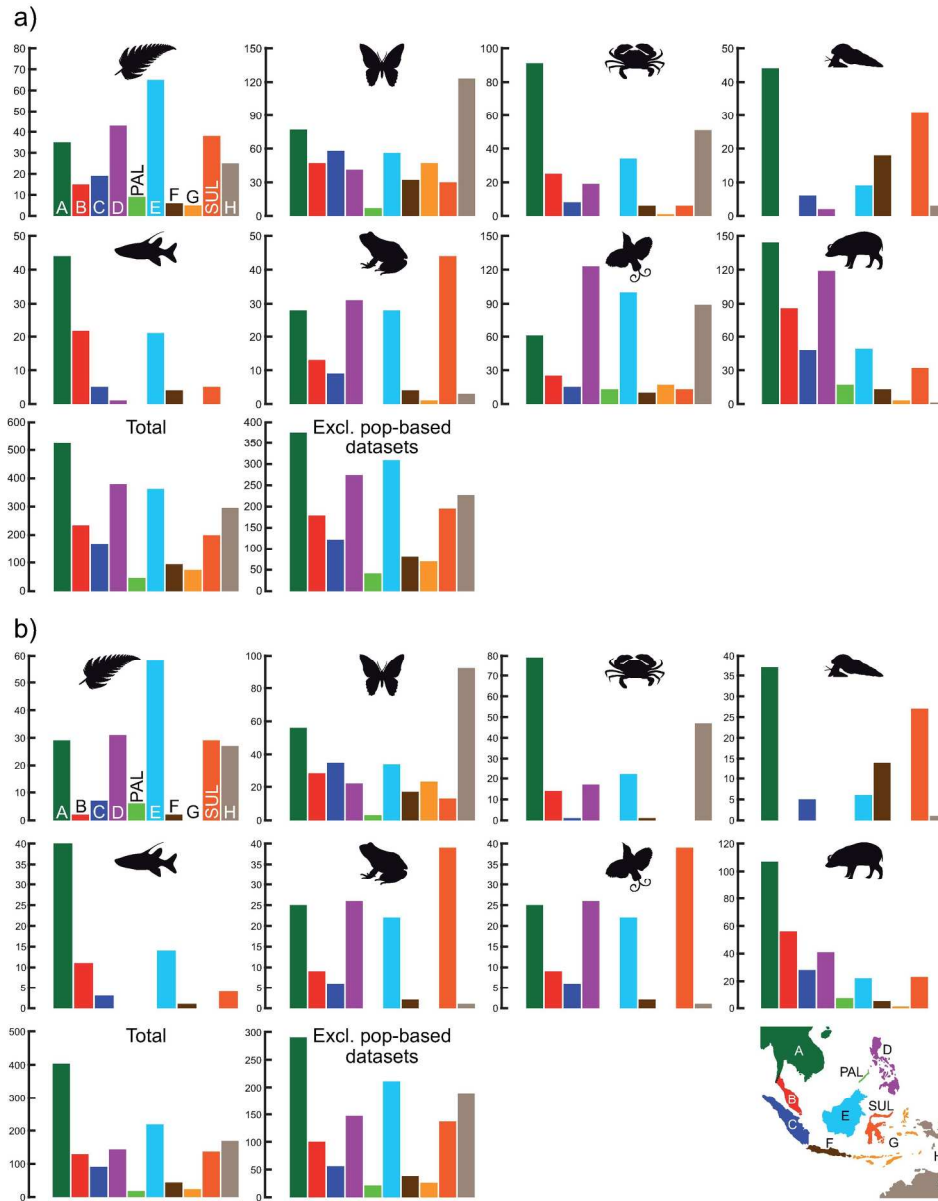


FIGURE 5. Distribution of lineages and in situ diversification events among all 61 datasets through time. a) Group-specific (plants, insects & spiders, freshwater crustaceans, freshwater molluscs, freshwater fishes, herpetofauna, birds, and mammals), total combined, and total combined excluding population-based number of lineages across all datasets from each pre-defined area. b) Group-specific (as above), total combined, and total combined excluding population-based number of in situ diversification events across all datasets from each pre-defined area. 191x245mm (300 x 300 DPI)

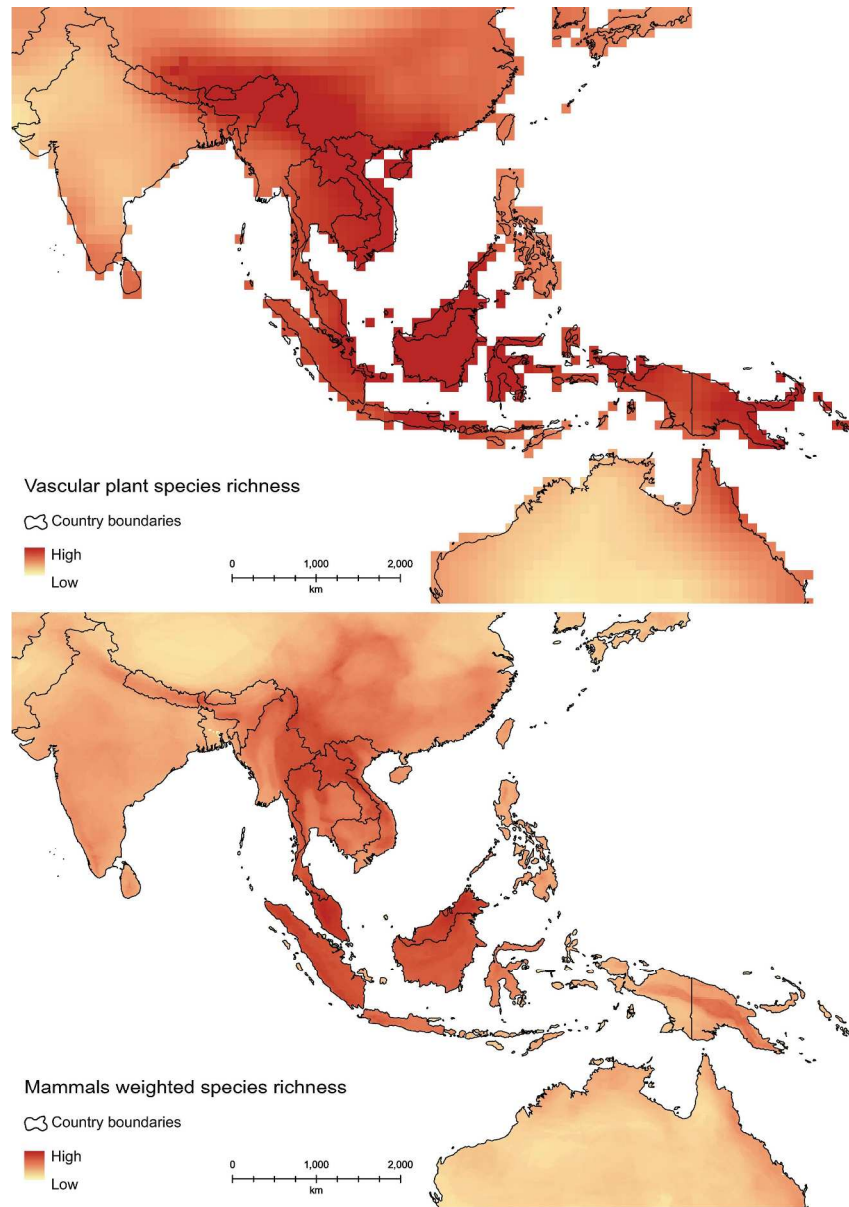


FIGURE 6. Extant Southeast Asian mammal and plant species richness calculated as the sum of species-specific distribution models. Borneo hosts the highest extant number of mammalian species ( $n=230$ ), and also the highest weighted species richness for both vascular plants and mammals (together with Thai-Malay Peninsula for top 10% of mammals) for SE Asia (see Material and Methods for details). Plant species richness data from Kreft and Jetz (2007), Copyright (2012) National Academy of Sciences, U.S.A. 212x302mm (300 x 300 DPI)

**Table 1.** List of datasets used for phylogenetic meta-analysis including information on taxa, generalized habitat, and molecular markers used in the original study.

| Dataset <sup>a</sup> | Common name         | Family           | Genus [generalized habitat] <sup>b</sup> | Genetic marker    | Taxonomic coverage <sup>f</sup>                | nTaxa (ingroup) <sup>g</sup> | BEAST/Lagrange Performed <sup>h</sup> | Source Ref <sup>f</sup> |
|----------------------|---------------------|------------------|--|-------------------|--|------------------------------|---------------------------------------|-------------------------|
| 01                   | plants              | Meliaceae        | <i>Aglaia</i> [1]                        | ITS               | 82 spec (all genera)                           | 42                           | this study                            | 42                      |
| 02                   | plants              | Araceae          | <i>Alocasia</i> [1]                      | cpDNA+nuDNA       | 71/113 species                                 | 73                           | *                                     | 46                      |
| 03                   | plants              | Begoniaceae      | <i>Begonia</i> [1]                       | cpDNA             | 92 spec (all families)                         | 66                           | #                                     | 61                      |
| 04                   | plants              | Gesneriaceae     | <i>Cyrtandra</i> [1]                     | ITS               | 26/300 species [A]                             | 30                           | this study                            | 2                       |
| 05                   | plants              | Ericaceae        | <i>Rhododendron</i> [1]                  | RPB2-d            | 46/300 species [B]                             | 46                           | this study                            | 20                      |
| 06                   | plants              | Rafflesiaceae    | several [1]                              | mtDNA+cpDNA+nuDNA | 80% spec (all genera)                          | 27                           | #                                     | 3                       |
| 07                   | mosquitoes          | Culicidae        | <i>Anopheles</i> [2]                     | mtDNA             | population-based dataset                       | 40                           | this study                            | 16                      |
| 08                   | butterflies         | Nymphalidae      | <i>Cethosia</i> [1]                      | mtDNA+nuDNA       | all species                                    | 42                           | #                                     | 43                      |
| 09                   | butterflies         | Nymphalidae      | <i>Charaxes</i> [1]                      | mtDNA+nuDNA       | all species                                    | 29                           | #                                     | 44                      |
| 10                   | butterflies         | Pieridae         | <i>Delias</i> [1]                        | mtDNA+nuDNA       | 30/44 species (Wallacea)                       | 131                          | #                                     | 69                      |
| 11                   | net-winged beetles  | Lycidae          | <i>Metriorrhynchus</i> [1]               | mtDNA             | all species                                    | 56                           | §                                     | 5                       |
| 12                   | giant wood spiders  | Tetragnathidae   | <i>Nephila</i> [1]                       | COI               | population-based dataset                       | 56                           | this study                            | 59                      |
| 13                   | cockroaches         | Blaberidae       | <i>Salganea</i> [1]                      | COII              | 22/50 species [C]                              | 36                           | this study                            | 33                      |
| 14                   | net-winged beetles  | Lycidae          | <i>Scarelus</i> [1]                      | mtDNA             | 18/32 species                                  | 18                           | *                                     | 35                      |
| 15                   | cockroaches         | Blaberidae       | several [1]                              | COII              | 21 species (9/10 genera)                       | 22                           | this study                            | 34                      |
| 16                   | mite harvestmen     | Stylocellidae    | several [2]                              | mtDNA+nuDNA       | 98/300 putative species (36 species described) | 95                           | this study / <sup>1</sup>             | 10                      |
| 17                   | fig wasps           | Sycophaginae     | several [1]                              | mtDNA+nuDNA       | 55 spec (all genera)                           | 55                           | #                                     | 12                      |
| 18                   | prawns              | Palaemonidae     | <i>Macrobrachium</i> [3]                 | COI               | population-based dataset                       | 93                           | this study / <sup>2</sup>             | 13                      |
| 19                   | prawns              | Palaemonidae     | <i>Macrobrachium</i> [3]                 | 16S rRNA          | 45/105 species                                 | 43                           | this study                            | 67                      |
| 20                   | crabs               | Gecarcinucidae   | several [3]                              | mtDNA+nuDNA       | 61 species (55% genera, 60% species)           | 61                           | §                                     | 26                      |
| 21                   | crabs               | Potamidae        | several [3]                              | 16S               | 72 species (51% genera, 14% species)           | 65                           | this study (AAR)                      | 57                      |
| 22                   | freshwater bivalves | Corbiculidae     | <i>Corbicula</i> [3]                     | COI               | 7/c. 30 species [D]                            | 55                           | this study / <sup>3</sup>             | 64                      |
| 23                   | freshwater snails   | Pachychilidae    | several [3]                              | 16S rRNA          | 21/c. 140 species (all genera)                 | 21                           | this study                            | 28                      |
| 24                   | freshwater snails   | Pachychilidae    | <i>Sulcospira</i> [3]                    | 16S rRNA          | 57/c. 140 species (all genera)                 | 129                          | this study / <sup>3</sup>             | 27                      |
| 25                   | cyprinids           | Cyprinidae       | <i>Barbodes</i> [3]                      | CR                | population-based dataset                       | 19                           | this study                            | 36                      |
| 26                   | ricefishes          | Adrianichthyidae | <i>Oryzias</i> [3]                       | mtDNA             | 13/24 species [E]                              | 23                           | this study                            | 60                      |
| 27                   | cobitids            | Cobitidae        | <i>Pangio</i> [3]                        | Cyt b             | 18/32 species                                  | 77                           | this study                            | 6                       |
| 28                   | fanged frogs        | Dicroglossidae   | <i>Limnonectes</i> [1]                   | mtDNA             | 45/55 species [F]                              | 78                           | this study                            | 19                      |
| 29                   | frogs               | Ranidae          | <i>Rana</i> [2]                          | ND3               | 14/? species (several cryptic species)         | 15                           | this study                            | 58                      |
| 30                   | frogs               | Ranidae          | <i>Rana</i> [1]                          | mtDNA             | 10/? species (several cryptic species)         | 54                           | this study                            | 21                      |
| 31                   | water snakes        | Homalopsidae     | several [3]                              | Cyt b             | 20/34 species                                  | 24                           | this study                            | 1                       |
| 32                   | spiderhunters       | Nectariniidae    | <i>Arachnothera</i> [1]                  | mtDNA             | 10/10 species                                  | 46                           | this study                            | 41                      |
| 33                   | bush warblers       | Cettiidae        | <i>Cettia</i> [1]                        | Cyt b             | 4/4 species (13/28 subspecies)                 | 12                           | this study                            | 48                      |
| 34                   | kingfishers         | Alcedinidae      | <i>Ceyx</i> [1]                          | ND2               | 4/6 species [G]                                | 19                           | this study                            | 30                      |
| 35                   | robins              | Turdidae         | <i>Copsychus</i> [1]                     | mtDNA             | 9/19 subspecies                                | 33                           | this study                            | 56                      |
| 36                   | flowerpeckers       | Dicaeidae        | <i>Dicaeum</i> [1]                       | ND2               | population-based dataset                       | 16                           | this study                            | 31                      |
| 37                   | forktails           | Muscicapidae     | <i>Enicurus</i> [1]                      | mtDNA             | population-based dataset                       | 15                           | this study                            | 40                      |

|    |                  |                 |                         |             |                                       |     |                           |        |
|----|------------------|-----------------|-------------------------|-------------|---------------------------------------|-----|---------------------------|--------|
| 38 | flycatchers      | Muscicapidae    | <i>Ficedula</i> [1]     | Cyt b       | 25/25 species                         | 27  | this study / <sup>4</sup> | 49     |
| 39 | cuckooshrikes    | Campephagidae   | <i>Lalage</i> [1]       | ND2         | population-based dataset              | 12  | this study                | 31     |
| 40 | spiderhunters    | Nectariniidae   | <i>Nectarinia</i> [1]   | ND2         | population-based dataset              | 30  | this study                | 31     |
| 41 | leaf-warblers    | Phylloscopidae  | <i>Phylloscopus</i> [2] | ND2         | population-based dataset              | 19  | this study                | 24     |
| 42 | bulbuls          | Pycnonotidae    | <i>Pycnonotus</i> [1]   | ND2         | population-based dataset              | 46  | this study                | 31     |
| 43 | fantails         | Rhipiduridae    | <i>Rhipidura</i> [1]    | ND2         | population-based dataset              | 50  | this study                | 31     |
| 44 | megapodes        | Megapodiidae    | several [1]             | ND2         | 15/22 species                         | 23  | this study                | 4      |
| 45 | white-eyes       | Zosteropidae    | several [1]             | mtDNA       | 42/c. 80 species ('clade B')          | 57  | this study                | 39     |
| 46 | whistlers        | Pachycephalidae | several [1]             | ND2         | 35/49 species                         | 39  | *                         | 25     |
| 47 | bulbuls          | Pycnonotidae    | several [1]             | mtDNA       | 43/> 130 species [H]                  | 46  | this study                | 47     |
| 48 | shrews           | Soricidae       | <i>Crocidura</i> [1]    | Cyt b       | 27/27 species (Malay Archipelago) [I] | 47  | §                         | 18     |
| 49 | wild dogs        | Canidae         | <i>Cuon</i> [2]         | CR          | population-based dataset              | 19  | this study                | 22     |
| 50 | fruit bats       | Pteropodidae    | <i>Cynopterus</i> [1]   | Cyt b       | 5/7 species [J]                       | 140 | this study                | 9 + 53 |
| 51 | elephants        | Elaphantidae    | <i>Elaphas</i> [2]      | mtDNA       | population-based dataset              | 32  | §                         | 63     |
| 52 | bats             | Hipposideridae  | <i>Hipposideros</i> [2] | ND2         | 21/70 species                         | 57  | this study                | 45     |
| 53 | macaques         | Cercopithecidae | <i>Macaca</i> [1]       | mtDNA+nuDNA | 5/20-22 species [K]                   | 23  | this study                | 62     |
| 54 | macaques         | Cercopithecidae | <i>Macaca</i> [1]       | mtDNA       | 15/20-22 species [K]                  | 17  | this study                | 68     |
| 55 | clouded leopards | Felidae         | <i>Neofelis</i> [1]     | mtDNA       | 2/2 species                           | 7   | this study                | 8      |
| 56 | tigers           | Felidae         | <i>Panthera</i> [1]     | mtDNA       | population-based dataset              | 25  | this study                | 32     |
| 57 | palm civets      | Viverridae      | <i>Paradoxurus</i> [1]  | mtDNA       | population-based dataset              | 86  | §                         | 50     |
| 58 | rodents          | Muridae         | several [2]             | mtDNA+nuDNA | 36 species (4/22 genera)              | 36  | this study / <sup>4</sup> | 23     |
| 59 | squirrels        | Sciuridae       | several [2]             | mtDNA       | 15 species (all genera)               | 15  | this study / <sup>4</sup> | 37     |
| 60 | tree squirrels   | Sciuridae       | <i>Sundascirus</i> [1]  | Cyt b       | 14/15 species                         | 29  | §                         | 15     |
| 61 | treeshrews       | Tupaiaidae      | <i>Tupaia</i> [1]       | mtDNA       | 20/20 species                         | 18  | §                         | 54     |

<sup>a</sup> Datasets showing 'Out-of-Borneo' signatures are marked grey.

<sup>b</sup> Generalized habitat type: 1 = rainforest, 2 = other habitat(s), 3 = freshwater.

<sup>c</sup> Calibration (F) = fossil calibration, calibration (G) = geological calibration, calibration (F+G) = fossil and geological calibration, calibration (I) = indirect calibration points from other study.

<sup>d</sup> Number of haplotypes/lineages/species according to source publication.

<sup>e</sup> Root height inferred from BEAST analyses; numbers in square brackets denote root ages directly obtained from source publication.

<sup>f</sup> Taxonomic coverage according to source authors; e.g., 71/113 species = 71 of 113 currently described species used in analysis; population-based datasets, see Table 3 for further details.

<sup>g</sup> Number of haplotypes (see source for information on lineages/species).

<sup>h</sup> AAR/Lagrange information: \* source publication uses Lagrange, DIVA, and MrBayes for AAR, respectively – data used for statistics and colonization routes in Fig. 4 and Fig. S14-15; # source publication performed AAR, but some defined areas summarize several islands (e.g., "Sundaland"); § AAR not performed in source publication; \$ AAR not performed in source publication but colonization routes inferred from topology; <sup>1</sup> not all terminal colonizations shown in Fig. 4 and Fig. S14 and S15, i.e. single lineages from Indochina, Thai-Malay Peninsula, Sumatra, and Borneo with uncertain ancestral area; <sup>2</sup> Lagrange not required, but topology suggests colonization from East of Wallacea to Sumatra in the Plio-Pleistocene (shown in Fig. 4 and Fig. S14), the remainder shows unresolved colonization routes between Indochina, Thai-Malay Peninsula, Sumatra, and Borneo; <sup>3</sup> Lagrange not feasible due to several unresolved basal relationships; <sup>4</sup> Lagrange not feasible due to several ambiguous distribution areas in certain taxa.

<sup>i</sup> Source reference: see Supplementary Information for source references.

**Table 2.** List of datasets used for phylogenetic meta-analysis (plus ancestral area reconstructions) including information on taxa, generalized habitat, substitution rates/fossil calibration applied by (source) authors, and results inferred from BEAST analyses (root height).

| Dataset | Family           | Genus                  | Substitution model used for analyses <sup>a</sup><br>part1/part2/part3 | jModelTest selection<br>AIC, BIC <sup>a</sup><br>part1/part2/part3 | Rates (%/Ma)/<br>dates used for<br>BEAST <sup>b</sup> | Root height (Ma):<br>mean (95% HPD) <sup>c</sup> | Source<br>ref <sup>d</sup> | Rates/dates<br>source<br>ref <sup>d</sup> |
|---------|------------------|------------------------|--|--|---|--|----------------------------|---|
| 01      | Meliaceae        | <i>Aglaia</i>          | G+G  | G+G, G+G   | calibration (F)                                       | 108.3 (70.8, 156.7)                              | 42                         | see source                                |
| 02      | Araceae          | <i>Alocasia</i>        |  |  | calibration (F)                                       | [47.5 (47.0, 48.0)]                              | 46                         | see source                                |
| 03      | Begoniaceae      | <i>Begonia</i>         |  |  | calibration (F)                                       | [22.3 (N.A., N.A.)]                              | 61                         | see source                                |
| 04      | Gesneriaceae     | <i>Cyrtandra</i>       | H+G*   | G+G   K80+G  | 0.1   | 167.9 (107.6, 245.2)                             | 2                          | see source                                |
| 05      | Ericaceae        | <i>Rhododendron</i>    | G+G  | G+G, G+G   | calibration (F)                                       | 59.9 (58.0, 61.9)                                | 20                         | 38  |
| 06      | Rafflesiaceae    | several                |  |  | calibration (I)                                       | [96.1 (83.6, 110.7)]                             | 3                          | see source                                |
| 07      | Culicidae        | <i>Anopheles</i>       | G+G/H+I+G  | G+G/H+I+G, G+G/H+I   | 2.3   | 1.5 (0.9, 2.2)                                   | 16                         | 7   |
| 08      | Nymphalidae      | <i>Cethosia</i>        |  |  | calibration (F)                                       | [67.0 (57.0, 82.0)]                              | 43                         | see source                                |
| 09      | Nymphalidae      | <i>Charaxes</i>        |  |  | calibration (F)                                       | [41.5 (38.5, 44.0)]                              | 44                         | see source                                |
| 10      | Pieridae         | <i>Delias</i>          |  |  | calibration (I)                                       | [N.A.]   | 69                         | see source                                |
| 11      | Lycidae          | <i>Metriorrhynchus</i> |  |  | 2.3   | [N.A.]   | 5                          | see source                                |
| 12      | Tetragrathidae   | <i>Nephila</i>         | H+G  | H+G, H+G   | 2.3   | 1.6 (0.9, 2.4)                                   | 59                         | 7   |
| 13      | Blaberidae       | <i>Salganea</i>        | \$   |  | 3.8   | 12.6 (9.4, 16.1)                                 | 33                         | 34  |
| 14      | Lycidae          | <i>Scarelus</i>        |  |  | 2.3   | [N.A.]   | 35                         | see source                                |
| 15      | Blaberidae       | several                | H+I+G  | H+I+G, H+I+G   | 3.8   | 17.2 (10.1, 24.9)                                | 34                         | see source                                |
| 16      | Stylocellidae    | several                | \$   |  | calibration (F)                                       | 425.0 (423.0, 427.0)                             | 10                         | see source                                |
| 17      | Sycophaginae     | several                |  |  | calibration (F+G)                                     | [48.2]   | 12                         | see source                                |
| 18      | Palaemonidae     | <i>Macrobrachium</i>   | H+G  | H+G, H+G   | 1.485   | 3.5 (2.2, 4.8)                                   | 13                         | 66  |
| 19      | Palaemonidae     | <i>Macrobrachium</i>   | H+G  | H+G, H+G   | 0.745   | 44.3 (24.3, 65.9)                                | 67                         | 14  |
| 20      | Gecarcinucidae   | several                |  |  | calibration (F)                                       | 54.5 (35.0, 76.6)                                | 26                         | see source                                |
| 21      | Potamidae        | several                |  |  | calibration (F+G)                                     | [N.A.]   | 57                         | see source                                |
| 22      | Corbiculidae     | <i>Corbicula</i>       | \$   |  | 1.48  | 1.9 (1.1, 2.9)                                   | 64                         | 66  |
| 23      | Pachychilidae    | several                | \$   |  | 1.0   | 78.7 (53.3, 106.8)                               | 28                         | see source                                |
| 24      | Pachychilidae    | <i>Sulcospira</i>      | G+G <sup>#</sup>   | H+G, H+G   | 1.0   | 13.2 (8.9, 18.0)                                 | 27                         | 28  |
| 25      | Cyprinidae       | <i>Barbodes</i>        | H+G  | H+G, H+G   | 3.6   | 9.3 (2.3, 20.7)                                  | 36                         | see source                                |
| 26      | Adrianichthyidae | <i>Oryzias</i>         | \$   |  | 2.5   | 7.7 (5.3, 10.4)                                  | 60                         | 17  |
| 27      | Cobitidae        | <i>Pangio</i>          | G+I+G  | G+I+G, H+I+G   | 0.68  | 48.0 (32.0, 66.4)                                | 6                          | see source                                |
| 28      | Dicroglossidae   | <i>Limnonectes</i>     | \$   |  | 1.48  | 22.8 (17.0, 29.5)                                | 19                         | 11  |
| 29      | Ranidae          | <i>Rana</i>            | G+G  | G+G, G+G   | 2.4   | 13.7 (9.3, 18.5)                                 | 58                         | 52  |
| 30      | Ranidae          | <i>Rana</i>            | H+I+G*   | G+I+G, G+I+G   | 2.4   | 11.4 (7.7, 15.3)                                 | 21                         | 52  |
| 31      | Homalopsidae     | several                | \$   |  | calibration (F)                                       | 41.9 (29.6, 55.1)                                | 1                          | see source                                |
| 32      | Nectariniidae    | <i>Arachnothera</i>    | H+I+G/H+I+G*   | G+I+G/G+I+G, G+I+G/H+I+G   | 2.1   | 15.7 (10.8, 20.6)                                | 41                         | 65  |
| 33      | Cettiidae        | <i>Cettia</i>          | H+G  | H+G, H+G   | 2.1   | 11.5 (5.5, 18.0)                                 | 48                         | 65  |
| 34      | Alcedinidae      | <i>Ceyx</i>            | G+G  | G+G, H+G   | 2.1   | 7.2 (5.0, 9.5)                                   | 30                         | 65  |
| 35      | Turdidae         | <i>Copsychus</i>       | H+I+G/H+I*   | G+I+G/G+I, H+G/G+I   | 2.1   | 1.8 (1.0, 2.7)                                   | 56                         | 65  |
| 36      | Dicaeidae        | <i>Dicaeum</i>         | G+G  | G+G, G+G   | 2.1   | 7.9 (4.5, 11.4)                                  | 31                         | 65  |
| 37      | Muscicapidae     | <i>Enicurus</i>        | S+I+G/H+G  | S+I+G/H+G, S+G/H+G   | 2.1   | 7.1 (4.4, 9.7)                                   | 40                         | 65  |
| 38      | Muscicapidae     | <i>Ficedula</i>        | G+G  | G+G, H+G   | 2.1   | 7.7 (5.4, 10.6)                                  | 49                         | 65  |
| 39      | Campephagidae    | <i>Lalage</i>          | G+G  | G+G, H+G   | 2.1   | 7.7 (2.9, 12.1)                                  | 31                         | 65  |
| 40      | Nectariniidae    | <i>Nectarinia</i>      | G+G  | G+G, G+G   | 2.1   | 6.7 (4.9, 8.5)                                   | 31                         | 65  |
| 41      | Phylloscopidae   | <i>Phylloscopus</i>    | H+I+G  | H+I+G, H+G   | 2.1   | 7.4 (3.8, 11.1)                                  | 24                         | 65  |
| 42      | Pycnonotidae     | <i>Pycnonotus</i>      | H+I+G*   | G+I+G, H+G   | 2.1   | 8.4 (4.4, 12.8)                                  | 31                         | 65  |

|    |                 |                     |             |                          |                 |                     |        |            |
|----|-----------------|---------------------|-------------|--------------------------|-----------------|---------------------|--------|------------|
| 43 | Rhipiduridae    | <i>Rhipidura</i>    | H+I+G*      | G+I+G, G+I+G             | 2.1             | 9.6 (2.3, 24.8)     | 31     | 65         |
| 44 | Megapodiidae    | several             | \$          |                          | 1.79            | 22.6 (14.8, 30.6)   | 4      | 51         |
| 45 | Zosteropidae    | several             | G+I+G/H+I+G | G+I+G/H+I+G, G+I+G/H+I+G | 2.1             | 22.7 (19.3, 26.1)   | 39     | 65         |
| 46 | Pachycephalidae | several             |             |                          | 2.1             | [7.2 (4.1, 10.9)]   | 25     | 65         |
| 47 | Pycnonotidae    | several             | G+I+G/H+G   | G+I+G/H+G, G+I+G/H+G     | 2.1             | 20.1 (16.5, 23.7)   | 47     | 65         |
| 48 | Soricidae       | <i>Crociodura</i>   |             |                          | calibration (F) | [8.0 (7.2, 8.9)]    | 18     | see source |
| 49 | Canidae         | <i>Cuon</i>         | H+G         | H+G, H+G                 | 5.48            | 1.4 (0.6, 2.5)      | 22     | 29         |
| 50 | Pteropidae      | <i>Cynopterus</i>   | H+I+G*      | G+I+G, G+I+G             | 4.7             | 1.3 (0.8, 1.7)      | 9 + 53 | 55         |
| 51 | Elaphantidae    | <i>Elaphas</i>      |             |                          | calibration (F) | 7.4 (7.3, 8.0)      | 63     | see source |
| 52 | Hipposideridae  | <i>Hipposideros</i> | G+G         | G+G, G+G                 | 4.7             | 8.2 (5.3, 11.1)     | 45     | 55         |
| 53 | Cercopithecidae | <i>Macaca</i>       | G+G/H+G/H   | G+G/H+G/H, JC/H+G/K80    | calibration (F) | 7.3 (4.6, 10.3)     | 62     | see source |
| 54 | Cercopithecidae | <i>Macaca</i>       | H+G/H+G     | H+G/H+G, H+G/H+G         | calibration (F) | 9.9 (7.1, 12.8)     | 68     | see source |
| 55 | Felidae         | <i>Neofelis</i>     | H+G/H+G     | H+G/H+G, H/H+G           | calibration (F) | 7.1 (5.4, 9.0)      | 8      | see source |
| 56 | Felidae         | <i>Panthera</i>     | H+I+G*      | G+I+G, H+G               | 1.53            | 1.6 (0.8, 2.8)      | 32     | see source |
| 57 | Viverridae      | <i>Paradoxurus</i>  |             |                          | calibration (F) | [N.A.]              | 50     | see source |
| 58 | Muridae         | several             | G+I+G       | G+I+G, G+I+G             | calibration (F) | 140.4 (99.5, 185.8) | 23     | see source |
| 59 | Sciuridae       | several             | \$          |                          | calibration (F) | 39.0 (34.5, 45.2)   | 37     | see source |
| 60 | Sciuridae       | <i>Sundascirus</i>  |             |                          | calibration (I) | [10.0]              | 15     | see source |
| 61 | Tupaiaidae      | <i>Tupaia</i>       |             |                          | calibration (F) | [83.5 (68.0, 93.0)] | 54     | see source |

<sup>a</sup> Substitution models: G = GTR, H=HKY, S=SYM; e.g., G+I = GTR+I, H+G = HKY+G, S+I+G = SYM+I+G; \$ = dataset re-used from Stelbrink et al. (2012); \* = less complex model used but also Gamma and Propinv parameters from AIC (jModelTest) due to low ESS values for parameters 'prior' and 'posterior' (see text for details); # model selected by AIC resulted in low basal branch supports and thus in a non-monophyly of the ingroup – therefore, GTR+G was used.

<sup>b</sup> Calibration (F) = fossil calibration, calibration (G) = geological calibration, calibration (F+G) = fossil and geological calibration, calibration (I) = indirect calibration points from other study.

<sup>c</sup> Root height inferred from BEAST analyses; numbers in square brackets denote root ages directly obtained from source publication.

<sup>d</sup> Source reference: see Supplementary Information for source references.



**Table 3.** Number of colonizations found between two areas (redundant ‘Out-of’ colonizations in a single dataset are counted once only). Area codes in square brackets denote areas defined *a priori*: A=Indochina, B=Thai-Malay Peninsula, C=Sumatra, D=Philippines, PAL=Palawan, E=Borneo, F=Java, G=Wallacea excluding Sulawesi, SUL=Sulawesi, H=East of Wallacea. Top 20 are highlighted in gray. Note that total number of emigrations/immigrations presented in Table S4 and tested statistically for significance are higher as they also include emigrations/immigrations to/from India/Sri Lanka and Japan (see Fig. 4).

| Area-pair:<br>source–<br>destination | pre-<br>Miocene | Miocene | Plio-<br>Pleistocene | Total | Area-pair:<br>source–<br>destination | pre-<br>Miocene | Miocene | Plio-<br>Pleistocene | Total |
|--------------------------------------|-----------------|---------|----------------------|-------|--------------------------------------|-----------------|---------|----------------------|-------|
| A–B                                  | 1               | 4       | 10                   | 15    | E–A                                  | 1               | 9       | 2                    | 12    |
| A–C                                  |                 | 3       | 7                    | 10    | E–B                                  | 4               | 6       | 11                   | 21    |
| A–D                                  |                 | 4       | 6                    | 10    | E–C                                  | 1               | 3       | 8                    | 12    |
| A–PAL                                |                 |         | 1                    | 1     | E–D                                  | 3               | 6       | 5                    | 14    |
| A–E                                  |                 | 6       | 8                    | 14    | E–PAL                                | 1               | 3       | 2                    | 6     |
| A–F                                  |                 | 2       | 4                    | 6     | E–F                                  |                 | 5       | 4                    | 9     |
| A–G                                  |                 | 1       | 1                    | 2     | E–G                                  |                 | 2       |                      | 2     |
| A–SUL                                |                 | 3       | 6                    | 9     | E–SUL                                | 1               | 3       | 2                    | 6     |
| A–H                                  |                 | 3       | 3                    | 6     | E–H                                  | 1               | 4       |                      | 5     |
| B–A                                  | 1               | 1       | 3                    | 5     | F–A                                  |                 | 1       | 1                    | 2     |
| B–C                                  | 2               | 1       | 3                    | 6     | F–B                                  |                 |         | 1                    | 1     |
| B–D                                  | 1               |         |                      | 1     | F–C                                  | 1               | 1       |                      | 2     |
| B–PAL                                |                 |         |                      | 0     | F–D                                  |                 |         |                      | 0     |
| B–E                                  | 1               | 1       | 3                    | 5     | F–PAL                                |                 |         |                      | 0     |
| B–F                                  |                 |         |                      | 0     | F–E                                  |                 |         |                      | 0     |
| B–G                                  |                 | 1       | 1                    | 2     | F–G                                  |                 | 1       |                      | 1     |
| B–SUL                                | 1               | 1       |                      | 2     | F–SUL                                |                 |         |                      | 0     |
| B–H                                  |                 |         |                      | 0     | F–H                                  | 1               |         |                      | 1     |
| C–A                                  |                 | 1       | 1                    | 2     | G–A                                  |                 | 1       |                      | 1     |
| C–B                                  | 1               | 2       | 1                    | 4     | G–B                                  |                 |         |                      | 0     |
| C–D                                  |                 |         | 2                    | 2     | G–C                                  |                 |         | 1                    | 1     |
| C–PAL                                |                 |         |                      | 0     | G–D                                  |                 |         | 2                    | 2     |
| C–E                                  |                 | 1       | 5                    | 6     | G–PAL                                |                 |         |                      | 0     |
| C–F                                  | 1               | 1       | 5                    | 7     | G–E                                  |                 |         | 1                    | 1     |
| C–G                                  |                 |         |                      | 0     | G–F                                  |                 |         | 1                    | 0     |
| C–SUL                                |                 | 1       |                      | 1     | G–SUL                                |                 | 1       |                      | 1     |
| C–H                                  |                 |         |                      | 0     | G–H                                  |                 |         | 4                    | 4     |
| D–A                                  |                 |         | 3                    | 3     | SUL–A                                |                 | 1       |                      | 1     |
| D–B                                  |                 |         | 1                    | 1     | SUL–B                                |                 | 1       |                      | 1     |
| D–C                                  |                 | 1       |                      | 1     | SUL–C                                |                 |         |                      | 0     |
| D–PAL                                |                 | 2       | 3                    | 5     | SUL–D                                |                 | 1       | 1                    | 2     |
| D–E                                  |                 | 1       | 2                    | 3     | SUL–PAL                              |                 | 1       |                      | 1     |
| D–F                                  |                 |         | 1                    | 1     | SUL–E                                |                 |         |                      | 0     |
| D–G                                  |                 | 1       |                      | 1     | SUL–F                                |                 | 1       |                      | 1     |
| D–SUL                                |                 | 3       | 3                    | 6     | SUL–G                                |                 |         |                      | 0     |
| D–H                                  |                 | 2       | 4                    | 6     | SUL–H                                | 1               | 1       | 2                    | 4     |
| PAL–A                                |                 |         |                      | 0     | H–A                                  |                 |         | 1                    | 1     |
| PAL–B                                |                 |         |                      | 0     | H–B                                  |                 |         |                      | 0     |
| PAL–C                                |                 |         |                      | 0     | H–C                                  |                 | 1       | 1                    | 2     |
| PAL–D                                | 1               |         |                      | 1     | H–D                                  |                 |         | 5                    | 5     |
| PAL–E                                |                 |         |                      | 0     | H–PAL                                |                 |         |                      | 0     |
| PAL–F                                |                 |         |                      | 0     | H–E                                  |                 |         |                      | 0     |
| PAL–G                                |                 |         |                      | 0     | H–F                                  |                 | 1       |                      | 1     |
| PAL–SUL                              |                 | 1       |                      | 1     | H–G                                  |                 | 1       | 2                    | 3     |
| PAL–H                                | 1               |         |                      | 1     | H–SUL                                | 1               | 3       | 2                    | 6     |

**Table 4.** Total numbers of lineages, *in situ* diversification events, and number of mammal species related to area size. Area codes denote biogeographic areas defined *a priori*: A=Indochina, B=Thai-Malay Peninsula, C=Sumatra, D=Philippines, PAL=Palawan, E=Borneo, F=Java, G=Wallacea excluding Sulawesi, SUL=Sulawesi, H=East of Wallacea.

| Biogeographic area | Area [km <sup>2</sup> ] <sup>a</sup> | N lineages <sup>b</sup> | Lineages / 1,000 km <sup>2</sup> | N <i>in situ</i> diversifications <sup>b</sup> | Diversifications / 1,000 km <sup>2</sup> | N Mammals species <sup>c</sup> | Mammals species / 1,000 km <sup>2</sup> | N emigrations | Emigration / 1,000 km <sup>2</sup> |
|--------------------|--------------------------------------|-------------------------|----------------------------------|--|--|--------------------------------|---|---------------|------------------------------------|
| A                  | 1,938,743                            | 376                     | 0.19                             | 290  | 0.15                                     | 453                            | 0.23                                    | 95            | 0.05                               |
| B                  | 131,598                              | 179                     | 1.36                             | 100  | 0.76                                     | 67                             | 0.51                                    | 25            | 0.19                               |
| C                  | 473,481                              | 122                     | 0.26                             | 56   | 0.12                                     | 97                             | 0.20                                    | 23            | 0.05                               |
| D                  | 285,350                              | 274                     | 0.96                             | 147  | 0.52                                     | 150                            | 0.53                                    | 29            | 0.10                               |
| PAL                | 14,650                               | 41                      | 2.80                             | 21   | 1.43                                     | 58*                            | 3.96                                    | 3             | 0.20                               |
| E                  | 743,330                              | 309                     | 0.42                             | 211  | 0.28                                     | 190                            | 0.26                                    | 89            | 0.12                               |
| F                  | 138,794                              | 80                      | 0.58                             | 38   | 0.27                                     | 59                             | 0.43                                    | 8             | 0.06                               |
| G                  | 142,090                              | 69                      | 0.49                             | 26   | 0.18                                     | 22                             | 0.15                                    | 14            | 0.10                               |
| SUL                | 174,600                              | 195                     | 1.12                             | 137  | 0.78                                     | 51                             | 0.29                                    | 15            | 0.09                               |
| H                  | 6,072,464                            | 227                     | 0.04                             | 189  | 0.03                                     | 101                            | 0.02                                    | 20            | <0.01                              |

<sup>a</sup> Present-day area as in Fig. 6.

<sup>b</sup> Total numbers of lineages and *in situ* diversification events excluding population-based datasets.

<sup>c</sup> Source: \*Esselstyn J.A., Widmann P., Heaney L.R. 2004. The mammals of Palawan Island, Philippines. Proc. Biol. Soc. Washingt. 117:271–302; otherwise: Global Mammal Assessment program: Rondinini C., Di Marco M., Chiozza F., Santulli G., Baisero D., Visconti P., Hoffmann M., Schipper J., Stuart S.N., Tognelli M.F., Amori G., Falcucci A., Maiorano L., Boitani L. 2011. Global habitat suitability models of terrestrial mammals. Philos. Trans. R. Soc. London B 366:2633–2641.