Diverge and Conquer: Phylogenomics of southern Wallacean forest skinks (Genus: *Sphenomorphus*) and their colonization of the Lesser Sunda Archipelago

Sean B. Reilly,[1,2] [iD] Benjamin R. Karin,[1] Alexander L. Stubbs,[1] Evy Arida,[3] Umilaela Arifin,[1,4] Hinrich Kaiser,[5,6] Ke Bi,[1,7] Amir Hamidy,[3] Djoko T. Iskandar,[8,9] and Jimmy A. McGuire[1]

[1]*Museum of Vertebrate Zoology and Department of Integrative Biology, University of California, Berkeley, CA 94720, USA*

  [2]*E-mail: sreilly@ucsc.edu*

[3]*Museum Zoologicum Bogoriense, Indonesian Institute of Sciences, Cibinong, Indonesia*

[4]*Zoology Museum Hamburg, Leibniz Institute for the Analysis of Biodiversity Change, Martin-Luther-King-Platz 3 20146, Hamburg, Germany*

[5]*Department of Biology, Victor Valley College, 18422 Bear Valley Road, Victorville, California 92395, USA*

[6]*Zoologisches Forschungsmuseum Alexander Koenig, Adenauerallee 160 53113, Bonn, Germany*

[7]*Computational Genomics Resource Laboratory, California Institute for Quantitative Biosciences, University of California, Berkeley, CA 94720, USA*

[8]*Basic Sciences Commission, Indonesian Academy of Sciences, 11, Jl. Medan Merdeka, Selatan Jakarta 10110, Indonesia*

[9]*School of Life Sciences and Technology, Institut Teknologi Bandung, 10 Jalan Ganesha, Bandung 40132, Indonesia*

The archipelagos of Wallacea extend between the Sunda and Sahul Shelves, serving as a semipermeable two-way filter influencing faunal exchange between Asia and Australo-Papua. Forest skinks (Genus *Sphenomorphus*) are widespread throughout southern Wallacea and exhibit complex clinal, ontogenetic, sexual, and seasonal morphological variation, rendering species delimitation difficult. We screened a mitochondrial marker for 245 *Sphenomorphus* specimens from this area to inform the selection of 104 samples from which we used targeted sequence capture to generate a dataset of 1154 nuclear genes (~1.8 Mb) plus complete mitochondrial genomes. Phylogenomic analyses recovered many deeply divergent lineages, three pairs of which are now sympatric, that began to diversify in the late Miocene shortly after the oldest islands are thought to have become emergent. We infer a complex and nonstepping-stone pattern of island colonization, with the group having originated in the Sunda Arc islands before using Sumba as a springboard for colonization of the Banda Arcs. Estimates of population structure and gene flow across the region suggest total isolation except between two Pleistocene Aggregate Island Complexes that become episodically land-bridged during glacial maxima. These historical processes have resulted in at least 11 *Sphenomorphus* species in the region, nine of which require formal description. This fine-scale geographic partitioning of undescribed species highlights the importance of utilizing comprehensive genomic studies for defining biodiversity hotspots to be considered for conservation protection.

Oceanic archipelagos are ideal systems for studying evolution and diversification because their remoteness produces enduring isolation of colonizing populations while intensifying the effects of genetic drift and new selective forces that promote speciation (Whittaker and Fernández-Palacios, 2007; Losos and Ricklefs 2009). Some of the most influential islands that helped shape our understanding of these processes include the Hawaiian and Galapagos archipelagos, both of which are quite isolated from the nearest continental landmass and were formed by the relatively simple sequential process of hotspot volcanogenesis (Whittaker and Fernández-Palacios, 2007). These archipelagos represent simplified natural experiments in evolution relative to mainland communities, thereby allowing for new insights into fundamental processes driving speciation. In contrast with these hotspot island chains are archipelagos or clustered sets of archipelagos, such as the Philippines or Wallacea, which reside closer to continental source populations and were generated by combinations of more complex tectonic processes, including volcanic activity, subduction-driven uplift, accretionary build-up, and extensional processes (Hall 2009). Because these archipelagos are closer to mainland sources, they are expected to have a higher frequency of successful colonization events, resulting in richer biotas shaped by more dynamic processes of species accumulation and sorting. Thus, investigations of these archipelagos offer the potential to provide new insights into complex historical biogeographic scenarios and the resulting effects on species formation (Jones and Kennedy 2008; Esselstyn et al. 2010; Brown et al. 2013; Jønsson et al. 2017; Ali and Heaney 2021).

Southern Wallacea is a vast area containing hundreds of oceanic islands, including both volcanic and nonvolcanic islands, ranging in age from ∼10–12 Ma to <1 Ma (Hall 2009, 2011). Four main geological components are present, including the Sunda Arc, Sumba Island, the Inner Banda Arc, and the Outer Banda Arc (Fig. 1). The Lesser Sunda Archipelago extends eastward from the Sunda Shelf before the Inner and Outer Banda Arcs curl first northward and then back toward the west (Hall, 2011). Despite recent phylogenetic and population genetic studies revealing substantial undescribed species-level diversity in the region (Reilly et al. 2019a; Reilly et al. 2022a; Reilly et al. 2022b; Blom et al. 2019; Grismer et al. 2021), it is nevertheless true that for many groups of terrestrial animals, the Lesser Sunda Islands are relatively species-poor when compared to the extremely species-rich Sunda and Sahul continental shelves. Whereas the islands comprising the Greater Sunda and Sahul Shelves have each periodically become land-bridged with one another and with adjacent mainland areas when sea levels dropped during glacial maxima (Simpson 1977; Monk et al. 1997), the islands of Wallacea remained separated from the unified shelves by a series of deep marine barriers. This, along with other factors such as island size, relatively young island ages, and more variable climates

(Van Welzen et al. 2011), may explain the relative paucity of species within Wallacea because every naturally occurring Wallacean lineage either arrived by "sweepstakes" overwater dispersal or diversified in situ following overwater colonization by an ancestral lineage. Notably, the biotic assemblages on islands in the interior of the archipelago reflect the biogeographic equivalent of "serial passaging" with the ancestors of these species each having successfully completed a series of overwater dispersal events to reach these islands. While the Wallacean fauna has been characterized as having a high dispersal ability, they appear to be at a competitive disadvantage in continental biota communities (Oliver et al. 2018; White et al. 2021).

Some of the larger volcanic islands (e.g., Sumbawa, Flores) may be composites formed when once-separated volcanic paleo-islands enlarged and became amalgamated over time (Muraoka et al. 2005). Other groups of islands are thought to have episodically merged, forming Pleistocene Aggregate Island Complexes (PAICs), such as Lombok + Sumbawa, and a large conglomerate known as "greater Flores", which is thought to have included Komodo, Rinca, Flores, Adonara, Solor, and Lembata. Finally, in other cases, the channels separating islands are of such great depth that it is highly unlikely that they were ever connected during recent intraglacials, although the dynamic nature of the tectonics of this region cannot preclude prior connections during earlier intraglacial low-stands (R. Hall pers. comm.). These sets of PAIC islands set up testable hypotheses regarding the relationships and levels of divergence between populations within and among PAICs.

The herpetofauna of southern and eastern Wallacea has been touted as a textbook example of a two-way filter zone, as evidenced by the opposing clinal patterns of Asian and Australo-Papuan species as one moves eastward or westward through the archipelago (Darlington 1957; Whittaker and Fernández-Palacios, 2007). Indeed, given the configuration of the archipelago, the stepping-stone model makes sense for recent arrivals (Reilly et al. 2017; Reilly et al. 2019b; Jones et al. 2019). However, given the tectonic complexity of the archipelago, this model appears to be an oversimplification when applied to taxa that colonized the archipelago early (such as in the late Miocene (∼5.3–12 Ma) before the Banda Arcs and Sumba had become emergent). Indeed, phylogenetically informed biogeographical analyses for several taxa suggest that more complex patterns of dispersal and vicariance are prevalent (Tänzler et al. 2016; Gwee et al. 2017; Morinaka et al. 2017; Pedersen et al. 2018; Reilly et al. 2019a; Reilly et al., 2021; Reilly et al. 2022a; Reilly et al. 2022b; Toussaint et al. 2020).

Skinks of the genus *Sphenomorphus* are a diverse assemblage that occurs throughout Southeast Asia, Wallacea, Australo-Papua, and the Solomon Islands. However, the genus has not dispersed beyond the Solomon Islands, as have some other skink
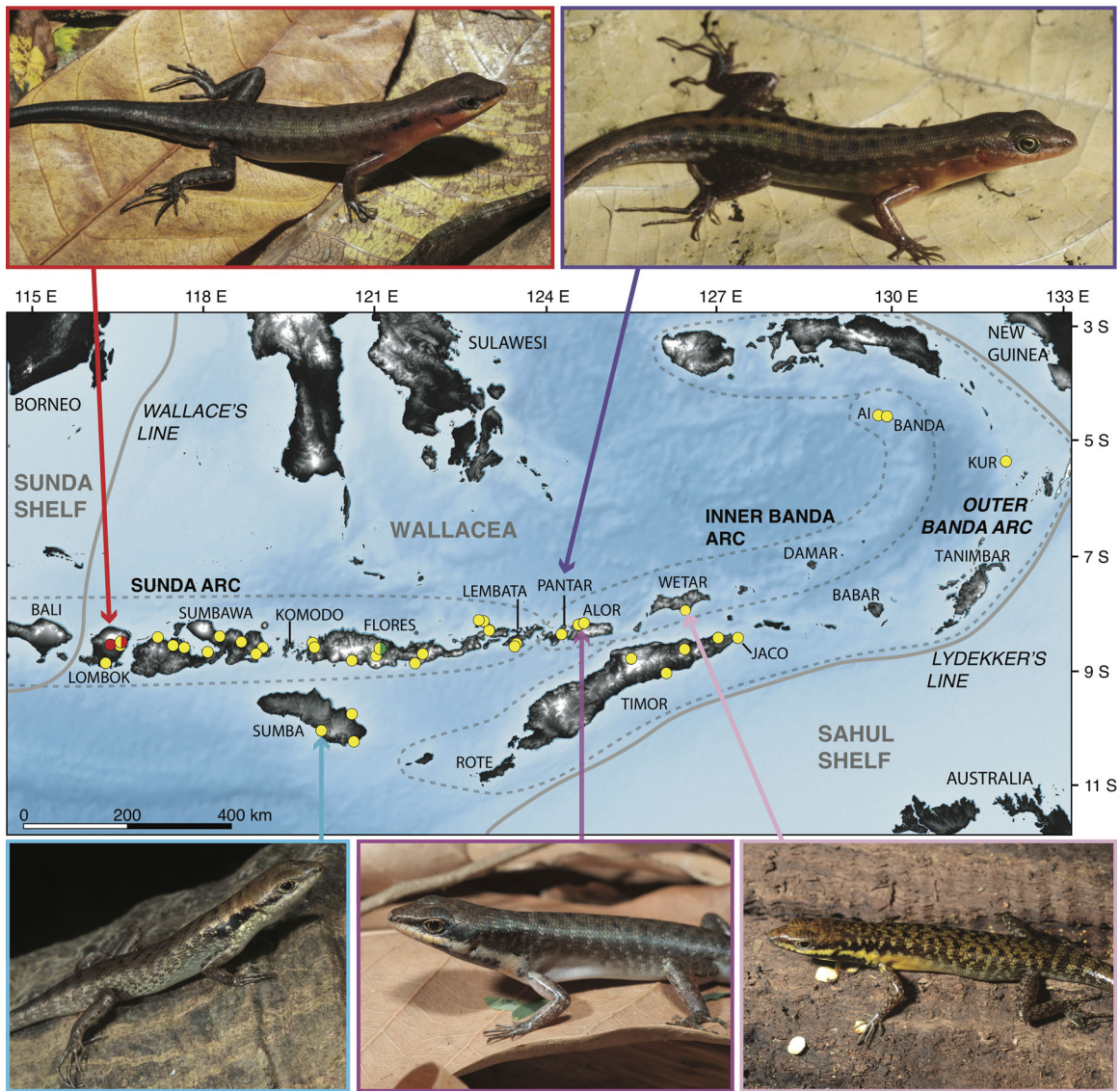
Figure 1. Map of the study area in southeastern Indonesia. Yellow circles represent exon-capture sampling localities for *Sphenomorphus melanopogon* complex samples, the half-green circle represents *S. striolatus*, and the red circles represent an undescribed species referred to as *S.* sp. *"Lombok"*. Localities with half colors represent sites where either sp. *"Lombok"* or *striolatus* were collected sympatrically with one of the *S. melanopogon* complex candidate species. Photos of *S. melanopogon* complex lizards from (clockwise from top left) Lombok, Pantar, Wetar, Alor, and Southwest Sumba.

genera (e.g., *Cryptoblepharus* [Blom et al. 2019], *Emoia* [Brown 1991], and *Lamprolepis* [Linkem et al. 2013]), suggesting that for skinks, *Sphenomorphus* may be considered to have moderate dispersal ability. The nominate species for the genus is *S. melanopogon* (Duméril & Bibron 1839), which occurs throughout southern Wallacea and on two small islands off the coast of southwestern Java, with the type locality being Timor (Duméril and Bibron 1839; Shea 2012; Karin et al. 2018; Reilly et al. 2020). Most of the Lesser Sunda Islands are inhabited by *S. melanopogon*, and this region makes up the majority of their range (Fig. 1). In the Lesser Sundas, a morphologically similar species, *S. strio-*

latus, occurs on Flores, but its relationship to *S. melanopogon* remains unknown. Commonly known as forest sinks, these diurnal lizards are moderately sized (sexual maturity reached at >45 mm snout-to-vent length; Shea 2012), with a variable color pattern composed primarily of shades of brown and black (Fig. 1). They occur from sea level to approximately 1200 m elevation and frequently occur in a variety of environments, including monsoon forests, dry creek beds, coastal moist forests, and ecotonal areas (Auffenberg 1980). They are primarily terrestrial but also occur in rocky areas and in trees up to 10 m in height (Auffenberg 1980). Until recently, *S. melanopogon* contained four subspecies based
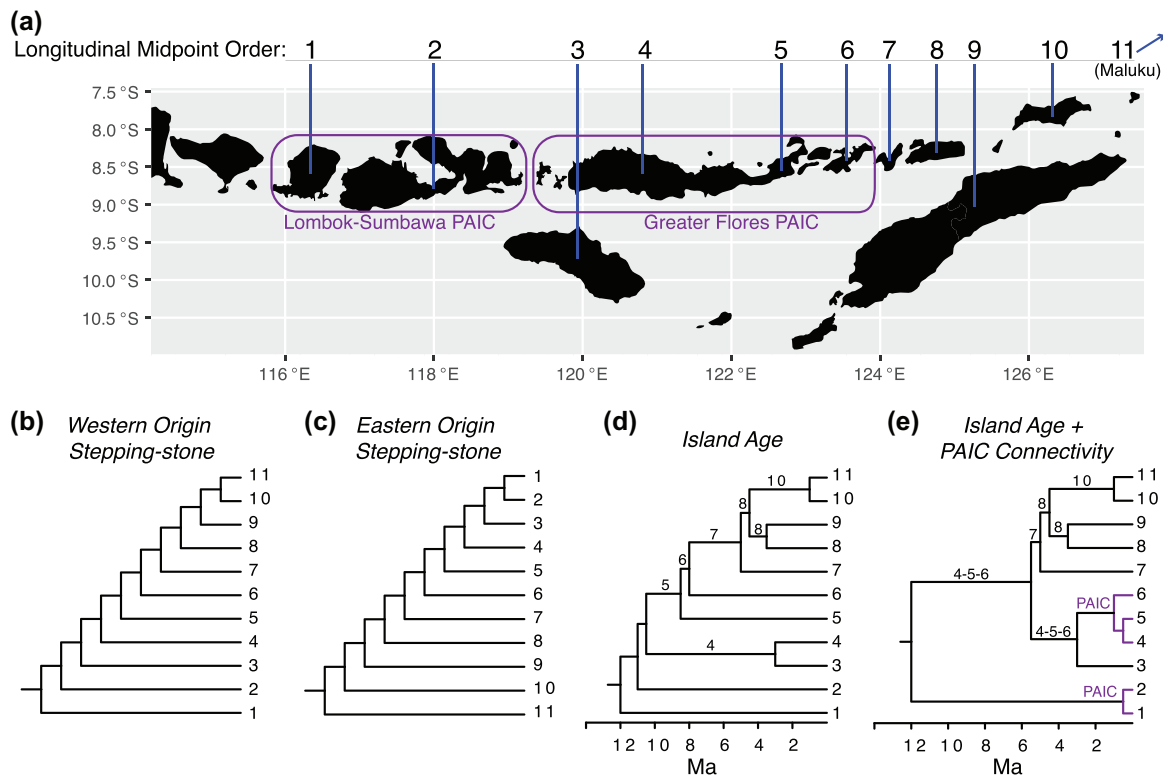
Figure 2. Phylogenetic hypotheses. (a) Simplified description of the region where islands are ordered from west to the east based on the longitude of their midpoint (Banda, Ai, and Kur are not shown but occur well to the east). The East Flores region is considered a separate region from West Flores based on previous studies. Tree topologies that would be consistent with various hypotheses regarding historical biogeography are presented. (b) A stepping-stone dispersal sequence from west to east. (c) A stepping-stone dispersal sequence from east to west. (d) A scenario where islands are colonized immediately after they become continuously emergent, and the nearest island serves as the source for later emerging islands. (E) A scenario mostly similar to the "Island Age" hypothesis but where islands that become periodically connected, or Pleistocene aggregate island complexes (PAICs), contain lineages that are each other's closest relatives. Numbers along internal branches in panels D-E represent ancestral ranges based on island age.

on divergent color patterns and geographical restriction to certain islands. Shea (2012) revised the species and synonymized all subspecies due to the presence of clinal variation in color patterns and overlapping scale characteristics. He further suggested (Shea 2012) that genetic data would be needed to address any further taxonomic division of *S. melanopogon*, with a special focus on the island of Flores. Based on our own field observations of morphological divergence and on preliminary mtDNA analyses, we hypothesize that *S. melanopogon*, which occurs across oceanic islands scattered over a >1000 km interval, comprises a species complex.

By ordering the Lesser Sunda Islands based on their midpoint longitudinal position (Fig. 2a), we are able to more clearly set some expectations based on various hypotheses. The *melanopogon* complex is thought to be sister to *S. meyeri* (Shea 2012), which occurs on the Sahul Shelf, although all other known close relatives, such as *S. indicus*, occur on the Sunda Shelf, leaving the source of southern Wallacean *Sphenomorphus* in question. If a topology is recovered with the more basal lineages in the west

and more derived lineages in the east, this would be consistent with a western origin stepping-stone hypothesis (Fig. 2b). However, if this pattern is reversed and the more basal lineages occur in the east while derived lineages occur in the west, this would support an eastern origin stepping-stone hypothesis (Fig. 2c). Alternatively, if islands are colonized immediately after they became emergent, regardless of the source, this would result in a time-calibrated tree where lineage ages were no older than the island age (Fig. 2d). An important aspect of this hypothesis involves recent dispersal from long-colonized islands to younger islands, which will appear to shorten the terminal branch length of the more ancient lineage(s). Finally, as mentioned above, some islands have become periodically land bridged, potentially allowing enough gene flow to prevent divergence and thus uniting the lineages on those island sets (Fig. 2e).

In this study, we screened the mitochondrial *ND4* gene for hundreds of *Sphenomorphus* samples from dozens of localities across 13 islands within southern Wallacea. This screening informed the selection of 104 samples from which we used targeted

sequence capture to generate a dataset of hundreds of orthologous nuclear loci and complete mitochondrial genomes.

These data were used to (1) reconstruct phylogenetic relationships, (2) assess population structure and identify any gene flow or introgression between lineages, (3) determine the level of divergence between lineages in relation to the speciation continuum, and (4) reconstruct the timing and sequence of island colonization. We discuss these findings in relation to the historical geology of the region and the biogeographical similarity to other codistributed species.

Materials and Methods

SAMPLE COLLECTION AND SCREENING

Sphenomorphus specimens were collected on the islands of Lombok, Sumbawa, Flores, Lembata, Pantar, Alor, Wetar, Timor, Jaco, Sumba, Kur, Banda, and Ai (Table S1). Specimens have been deposited in the Museum of Vertebrate Zoology at the University of California at Berkeley (MVZ), the Museum of Comparative Zoology at Harvard University (MCZ), the United States National Museum of Natural History, Smithsonian Institution (USNM), and the Museum Zoologicum Bogoriense (MZB).

DNA was extracted from liver or tail tissue using standard salt extraction techniques or the DNeasy kit (Qiagen). The *ND4* gene was sequenced for 245 samples (GenBank numbers MW291696–940) using standard PCR amplification with the primers ND4 and LEU (Arèvalo et al., 1994). Cleaned PCR products were sequenced on an ABI 3730 sequencer (Applied Biosystems, Foster City, California, USA), and raw sequence reads were combined in CODONCODE ALIGNER 3.5.2 (CodonCode Corporation, Dedham, Massachusetts, USA) and aligned with MUSCLE (Edgar, 2004). We estimated a maximum likelihood (ML) tree from the partitioned sequence alignment of 932 bp using IQTREE v2.1.1 (Chernomor et al. 2016; Minh et al. 2020) with automatic model selection according to BIC scores (Kalyaanamoorthy et al. 2017). Node support was assessed with 1000 ultrafast bootstrap replicates (UFBoot; Hoang et al. 2018) and 1000 SH-like approximate likelihood ratio tests (SH-aLRT), where UFBoot probabilities $\geq$95% and SH-aLRT values $\geq$80% indicate strong clade support. The single-locus species delimitation software ABGD (Puillandre et al. 2012) was used to identify lineages that are candidate species under the JC69 distance metric with the relative gap width parameter set to 1.5.

MARKER DEVELOPMENT AND EXON-CAPTURE DATA COLLECTION

Following Bi et al. (2012), we sequenced and compared transcriptomes from one *S. striolatus* and two *S. melanopogon* samples to identify orthologous loci (Table S2). We chose 1199 exons repre-

senting 1.09 Mb of sequence data from the two *S. melanopogon* transcriptomes ($\sim$1.6% sequence divergence) to serve as references for 3$\times$ tiling of 120 bp probes across each reference. These probes were implemented in an in-solution exon capture and enrichment experiment using the MYcroarray MYbaits kit (now Arbor Scientific). We created biotinylated bait libraries for 104 samples following Meyer and Kircher (2010) with slight modifications. These samples include *Tytthoscincus parvus, Sphenomorphus indicus, S. maculatus*, *S. meyeri* (Aru Island, Indonesia)*, S. striolatus* (Flores Island, Indonesia), seven samples of an undescribed species from Lombok (referred to as "sp. Lombok"), and 91 *S. melanopogon* complex samples from Lesser Sundas and Maluku Provinces (Table S1). The MYbaits reactions were performed following recommendations in the v2.3.1 manual with minor alterations. Postcapture reactions were amplified and pooled to be sequenced on one lane of an Illumina HiSeq2500 with 100 bp paired-end reads. Detailed methods concerning transcriptome sequencing, marker development, sample library preparation, the exon-capture experiment, and bioinformatics processing of the data can be found in the Supplementary Methods.

MITOGENOMICS

Although mitochondrial genes were not included in our probe design, we were able to obtain complete mitochondrial genomes as bycatch from the nontarget reads. The cleaned Illumina reads were mapped to the mitochondrial genome of *S. huanrenensis* (GenBank number NC030779), assembling contigs for each sample and aligning the assemblies in GENEIOUS PRIME version 2019.2. After removing the control region, tRNAs, and ribosomal RNA genes, 15,419 base pair coding gene alignments were obtained for 102 of the 104 samples, with two samples removed due to low coverage. This alignment was partitioned by gene and codon using BIC scores in PARTITIONFINDER2 (Lanfear et al. 2017), and a phylogenetic estimate was obtained using maximum likelihood inference with RAXML (Stamatakis 2014). Nodal support was assessed using 1000 nonparametric bootstrap pseudoreplicates. A time-calibrated Bayesian phylogeny was estimated using STARBEAST2 (Ogilvie et al. 2017) using the GTR model of sequence evolution, a Yule tree prior, and an uncorrelated log-normal relaxed clock with a rate of 0.02 mutations per site per 10^6 years, which matches the mean mitochondrial genome mutation rate for vertebrates (Allio et al. 2017). Two runs of 100 million generations were combined after a 10% burn-in was removed from each, and convergence was assessed by confirming that ESS values were greater than 200 for all parameters in TRACER version 1.7 (Rambaut and Drummond 2009). Levels of sequence divergence between major lineages or between biogeographically relevant populations were calculated using the

web server DIVEIN (Deng et al. 2010). The candidate species discovery software ABGD (Puillandre et al. 2012) was run under the JC69 model with the relative gap width parameter set to 1.0 to compare with our *ND4* ABGD analysis and to compare with phylogenomic results.

PHYLOGENOMICS

The partitioned alignment of all exon-capture sequence data was analyzed with the maximum likelihood method implemented in RAXML (Stamatakis 2014) under the GTRCAT model of sequence evolution, with nodal support assessed with 1000 bootstrap pseudoreplicates. The data were partitioned by exon plus flanking region. Species tree estimation was conducted using three approaches: (1) a summary coalescent approach using ASTRAL-III (Zhang et al. 2018), which estimates the species tree from gene tree input data by maximizing the number of shared induced quartets; (2) a Bayesian multispecies coalescent method using STARBEAST2 (Ogilvie et al. 2017); and (3) a full multispecies coalescent (MSC) approach as implemented in BPP version 4.2.9 (Yang and Rannala 2010; Rannala and Yang 2013).

For the ASTRAL-III analysis, we estimated individual gene trees for 1064/1154 genes (90 genes were not analyzed due to higher levels of missing data) with RAXML (Stamatakis 2014) under the GTR + I + G model of sequence evolution and used these gene trees as the input data. A priori species designations were based on biogeographically relevant clusters as identified in the concatenated maximum likelihood phylogenetic analysis. ASTRAL-III branch lengths are in coalescent units, and node support was assessed with local posterior probabilities computed on the resulting species tree (Sayyari and Mirarab 2016).

Because STARBEAST2 is a computationally intensive method for subgenomic datasets, we chose two samples from each major lineage along with our outgroups for a total of 36 samples. We selected 200 loci that contained sequences for all 36 samples in the 300–1500 bp size range to ease the computational burden. The final dataset contained a total of 203,373 bp for an average locus size of ∼1017 bp. A priori species designations were set based on well-supported clusters identified in the concatenated ML analysis. Four separate runs of 100 million generations sampled every 10,000 generations were carried out under the gamma site model, the JC69 mutation model, and a strict molecular clock linked between loci with a clock rate of 0.001 (Brandley et al. 2011; Blom et al. 2016; Allio et al. 2017). Stationarity was assessed by examining trace plots of parameters for each run to confirm convergence and to identify the proportion of burn-in required for each run. Once the burn-in samples were removed and runs were combined, we confirmed that ESS values for parameters were greater than 200 in TRACER version 1.7 (Rambaut and Drummond 2009).

We also employed BPP version 4.3.8 (Yang, 2015) to estimate a multispecies coalescent time tree for this assemblage. This version of BPP incorporates the multispecies coalescent with introgression (MSCi) model, which can account for both incomplete sorting and gene flow while estimating effective population sizes and time since divergence on a fixed species tree topology. For these analyses, we included the full 1154 sequence locus dataset for 101 individuals representing 15 a priori designated species (a much larger number of samples and loci than were feasible with STARBEAST2). All analyses were run for 1 million generations with a burnin of 16,000 generations and the following prior settings (theta = 3 0.002 e; tau = 3 0.002; and finetune = 3 0.003 0.002 0.00002 0.005 0.9 0.001 0.001). Two sets of analyses were undertaken. In the first set of runs, we employed the MSCi model on a fixed tree topology (obtained from the STARBEAST2 and ASTRAL analyses) with introgression bands that we selected based on potential signatures of introgression observed in STRUCTURE output. We considered three possible cases of bidirectional migration, including between (1) Lombok and Sumbawa, (2) West Flores and *S. striolatus*, and (3) West Flores and East Flores 2. For these analyses, the phi migration parameter prior was set to "1 1". Because these analyses suggested limited if any actual gene flow occurred between these lineages, we followed up with a full MSC time tree search with the MSC model (without an introgression parameter) using the same prior settings as outlined above with the exception of the excluded phi migration parameter.

As an additional metric of genetic divergence between lineages, we calculated genealogical divergence index values (*gdi*; Jackson et al. 2017). These values are on a scale of 0 to 1, with values <0.2 considered strong support for conspecificity, values >0.7 indicating strong support for distinctiveness at the species level, and intermediate values indicating ambiguity (Jackson et al. 2017; Leaché et al. 2019). We obtained *gdi* values using the divergence time (τ) and population size (θ) parameters estimated with the A00 function (see above) in the multispecies coalescent (MSC) program BPP version 4.3.8 (Yang and Rannala 2010; Rannala and Yang 2013). The *gdi* for each population is calculated by summarizing the entire retained posterior distribution using the equation from Jackson et al. (2017). Convergence for all BPP analyses was assessed by comparing the posterior distributions from two independent runs and by confirming that all parameter ESS values were greater than 200 in Tracer version 1.7 (Rambaut and Drummond 2009).

BIOGEOGRAPHICAL MODEL COMPARISON

Biogeographical models and dispersal modifiers were compared using the software BIOGEOBEARS (Matzke 2013; Matzke 2014) with the goal of testing different hypotheses regarding the sequence of island colonization (e.g., stepping-stone versus island

connectivity), as well as the timing (e.g., correlated with island age or not). The primary input of this analysis was the STAR-BEAST2 time-calibrated genomic phylogeny described above. Each sample was assigned to one of 11 regions: (1) Lombok, (2) Sumbawa, (3) West + Central Flores, (4) East Flores, (5) Lembata, (6) Pantar, (7) Alor, (8) Wetar, (9) Timor + Jaco, (10) Sumba, and (11) islands in Maluku Province (Ai, Banda, Kur). The program was run with four different dispersal modifiers: (1) an "Equal" model where each region has an equal dispersal probability to any other region; (2) a "Distance" model (Van Dam and Matzke 2016) where the dispersal probability between any two regions is scaled by the minimum distance between them (in km) as estimated using Google Earth (2020); (3) a "Stepping-Stone" model (Table S3) where dispersal is highly favored between adjacent regions (dispersal probability = 1) and discouraged between nonadjacent regions (dispersal probability = 0.00001); and (4) a "PAIC" model (Table S3), which is a modification of the Stepping-Stone model where the highest dispersal probabilities are between regions that are currently connected or become connected during glacial maxima (based on channel depths). These four dispersal modifiers were run under the biogeographical models DEC, DEC + J, DIVALIKE, DIVALIKE+J, BAYAREALIKE, and BAYAREALIKE+J. These 24 different models (four dispersal modifiers, six models) were compared using the log likelihood (LnL), Akaike information criteria (AIC), and corrected AIC (AICc) scores to determine the most likely models.

POPULATION STRUCTURE

To confirm the distinctiveness of lineages identified using phylogenetic analyses, we utilized a nonmodel-based principal component analysis (PCA) of genetic covariance (ADEGENET; Jombart 2008) and a model-based Bayesian clustering program (STRUCTURE version 2.3.2; Pritchard et al. 2000) to assign individuals to genetic clusters. These two methods are also useful for detecting hybrid individuals. For the PCA, a total of 171,751 informative SNPs (excluding singleton SNPs) served as the input to analyze four groups: (1) all *Sphenomorphus* from the Lesser Sundas and Maluku, (2) a "Sunda Arc clade" of *S. melanopogon*, (3) a "widespread clade" of *S. melanopogon*, and (4) an "eastern subclade" consisting of Alor, Wetar, Timor, and Jaco in the Lesser Sundas along with islands in Maluku Province. The results of the top two components were plotted against each other, and biogeographically cohesive groups of samples were highlighted. For clustering analyses, one informative SNP per locus served as the input. Analyses were run without specifying prior location under the admixture model with allele frequencies correlated. Ten replicates per preset number of populations (K) were run with 250,000 burn-in and 250,000 retained generations from K = 1–7 and 350,000 burn-in and 350,000 retained generations for K >7. The range of K explored for each group was based on hypotheses

of population structure based on phylogenomic and PCA results (*S. melanopogon* = 1–12; Sunda Arc clade = 1–6; Widespread clade = 1–10). The optimal number of K was estimated using the delta-K analysis (Evanno et al. 2005) implemented in the STRUCTURE HARVESTER (Earl and vonHoldt 2012). Ancestry bar plots for the optimal K and the largest K that returned biogeographically meaningful results were created with final runs utilizing a 1,000,000 burn-in and 1,000,000 retained generations.

To test for signatures of isolation-by-distance (IBD), which if confirmed could support a stepping-stone dispersal pattern, we conducted Mantel tests on matrices of Euclidean geographic distances (in km) and Edwards' genetic distances using the mantel.randtest function in ADEGENET. Significance was assessed with 100,000 permutations. Genetic distance was plotted against geographic distance for each sample pair, and local point density was calculated with two-dimensional kernel density estimation to produce a colored cloud over the points with the R package MASS (Ripley and Venables 2002). This approach was taken for three subsets of samples: (1) *S. melanopogon*, (2) the Sunda Arc clade, and (3) the widespread clade.

GENE FLOW AND INTROGRESSION

We employed two strategies to determine the level of genetic isolation between lineages that were identified using phylogenetic and population structure methods: (1) we used an isolation-with-migration program to estimate the levels of ongoing gene flow, and (2) we performed both four- and five-taxon tests to detect more ancient introgression between lineages. Estimates of population divergence times (τ) and migration rates (m) were obtained using the program G-PHOCS (Gronau et al. 2011), which is capable of dealing with genomic sequence data from unlinked neutrally evolving loci. The full dataset of introns plus exons (1154 loci) and even the full flanking dataset (1049 loci) proved computationally unviable, and tests on multiple sets of 500 loci produced nearly identical results, so one set of 500 flanking loci was chosen for all analyses. Each analysis was run for 1,000,000 generations and summarized after the removal of 100,000 generations of burn-in. All runs were checked in TRACER version 1.7 (Rambaut and Drummond 2009) to assess parameter convergence and confirm that all ESS values were greater than 200. A mutation rate (μ) of 1×10^{-9} mutations/site/year was used to convert parameter estimates to real-world values (Brandley et al. 2011; Blom et al. 2016; Allio et al. 2017). Priors for θ and τ were set to 1 and 10,000 for alpha and beta, respectively, and migration priors were set to 0.002 and 0.00001 for alpha and beta, respectively. We converted the output parameters into population divergence time in years (T) and migrants per generation (calculated independently of mutation rate) using equations found in the supplementary material of Gronau et al. (2011). A generation time of 1 year was used for *S. melanopogon* based on other

tropical skink species (Alcala 1966) as well as more distantly related species with similar life histories (Vitt and Cooper 1986).

We used COMPD (Mussmann et al. 2020) to perform the "ABBA-BABA" four-taxon test (Green et al. 2010; Durand et al. 2011) and the "Dfoil" five-taxon test (Pease and Hahn 2015), which detect statistically significant imbalances of discordant biallelic site patterns indicative of introgression (Martin et al. 2013; Eaton and Ree 2013). These methods have been shown to effectively detect introgression from targeted-sequence datasets (Pease and Hahn 2015; Lambert et al. 2019; Mussmann et al. 2020), although they cannot detect introgression between sister taxa. The exon-capture alignment of 1,801,789 bp was converted to biallelic site data with heterozygous sites included (Eaton et al. 2015). Individual tests were performed on one individual per taxon, and all possible combinations of samples from each population were tested. A population-wide Z score and p value were estimated from 1000 bootstrap replicates per test. A Z score >3 or <-3 and a p value <0.001 were the limits used to indicate significant introgression (Durand et al. 2011; Pease and Hahn 2015; Mussmann et al. 2020).

Results

SAMPLE SCREENING

The *ND4* phylogeny recovered a number of well-supported, geographically defined lineages, although the support for the relationships among those lineages varies (Figs. S1 and S2). Candidate species analyses suggest that the dataset contains 16 species (Fig. S2). An unexpected candidate species was identified from Lombok Island (referred to here as "sp. Lombok", one of two species inferred for this island). This unexpected lineage was found to be basal to a clade composed of *S. striolatus* + *S. melanopogon*, although this clade has weak support (UFBoot = 9.1; SH-aLRT = 82). Individual islands that are represented by single endemic candidate species include Sumbawa, Alor, and Pantar. Single islands that contain multiple candidate species include Lombok (2), Flores (6), Sumba (2), and Timor (2). Candidate species that occur on multiple islands include one from Wetar + Ai + Banda + Kur, one from Timor + Jaco, and one from Flores + Lembata.

EXON-CAPTURE DATA CHARACTERISTICS

A total of 1154 out of the original 1199 targeted loci passed the filtering stages. The total concatenated alignment consisting of both targeted exons and flanking introns was 1,801,789 base pairs of sequence data, averaging 1563 bp per locus. All 104 libraries were successfully sequenced and retained for analyses. The average coverage of the targeted regions was approximately 82X (e.g., an average of 82 unique reads aligned at any given site) after duplicate reads were removed, although individual sample

coverage ranged from ∼25X up to ∼140X (Fig. S3). The flanking regions had approximately 35X average coverage after duplicates were removed, with individual sample coverages ranging from ∼10X to ∼50X (Fig. S4). More than 500 of the 1154 alignments contained all 104 samples, and the average number of samples per alignment was 103 (Figs. S5B and D). The number of informative sites has a relatively linear relationship with the length of the loci with very few outlier loci (Fig. S5A), and there are on average ∼11.5% informative sites per alignment (Fig. S5F). There is no clear relationship between the alignment length and the percentage size of gaps in each alignment (Fig. S5C). The final length of the contigs ranged from 100 bp up to ∼4200 bp, and the percentage of missing data was ≤25% after the additional filtering step (Fig. S5E and G).

MITOGENOMICS

The partitioned Maximum Likelihood mitogenomic phylogeny (Fig. 3B) generally has better supported nodes than the *ND4* phylogeny, including a notable improvement for the sister taxon relationship of *S. striolatus* and *S. melanopogon* (BS = 96). Some newly recovered relationships, such as Alor + Timor, are not well supported (BS = 8). Some notable topology differences include the following: (1) within *S. melanopogon*, the Lombok lineage is found to be sister to Sumbawa instead of West Flores, (2) the East Sumba lineage is now recovered as sister to a Timor + Jaco + Alor + Wetar + Maluku clade, and (3) all samples from Timor + Jaco form a clade, although with low support (BS = 51). The candidate species analysis also differs from that using the *ND4* data by recovering a 17th candidate species (Banda + Ai). The candidate species delimitation within West Flores also differs from the *ND4* results. The uncorrected sequence divergence between the *Sphenomorphus* sp. "Lombok" and *S. melanopogon* lineages range from 17.7% to 18.8%, and the *S. striolatus* and *S. melanopogon* lineages range from 18.2% to 19.3% (Table S4). The levels of sequence divergence within *S. melanopogon* ranged from 2.7% (Kur/Wetar) to 16.5% (Sumbawa/East Sumba).

The time-calibrated Bayesian tree had moderate to high node support in all runs with 5/16 nodes returning posterior probability values <0.8, although it always returned a majority rule consensus tree with a topology identical to the ML tree (Fig. S6). Although mitochondrial divergence time estimates should be interpreted with caution, analysis of the mitochondrial dataset suggests that *Sphenomorphus* sp. Lombok diverged between ∼6.0 and 9.5 million years ago (mean = 7.75 million years ago) and that *S. striolatus* diverged from the *melanopogon* complex between 5.5 and 8.0 million years ago (mean = 6.77 million years ago). The Sunda Arc and widespread clades are estimated to have diverged ∼4.5–6.5 million years ago (mean = 5.63 million years ago). The youngest divergence times are between Wetar and Kur
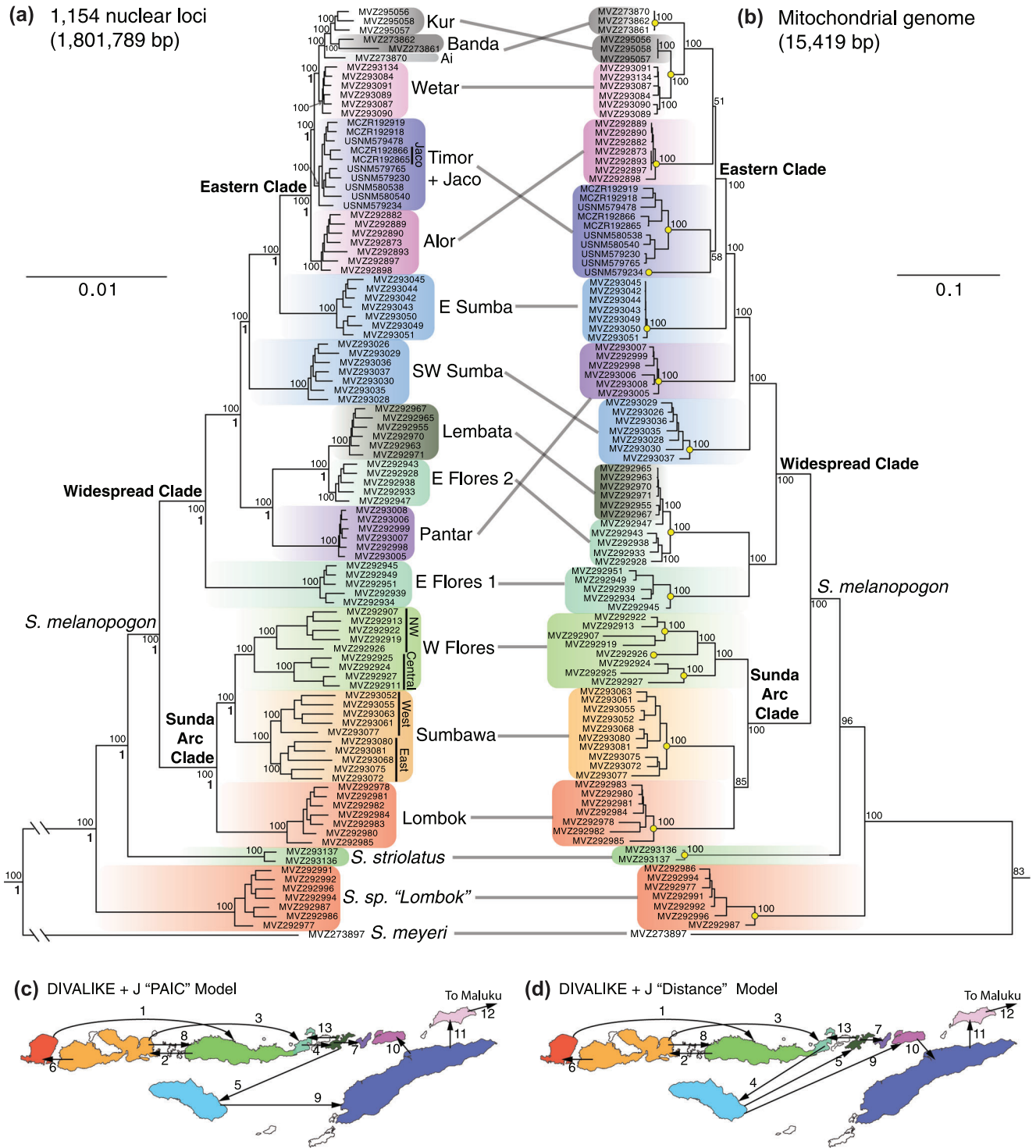
Figure 3. Maximum likelihood phylogenies for (a) the genomic dataset and (b) the mitochondrial genome. Numbers at nodes represent bootstrap support, and bold numbers in the genomic phylogeny represent ASTRAL local posterior probability. Yellow circles at mitochondrial tree nodes delimit candidate species. Trees are rooted by the outgroup samples of *Tytthoscincus parvus*, *Sphenomorphus indicus*, and *S. maculatus*. Best joint histories from the top supported biogeographical model (DIVALIKE + J) and dispersal modifiers (c = PAIC; d = Distance). Arrows represent dispersal events, with numbers indicating their temporal sequence. It should be noted that the probability of the best joint history is low given the uncertainty at many of the nodes (see Figure S8).
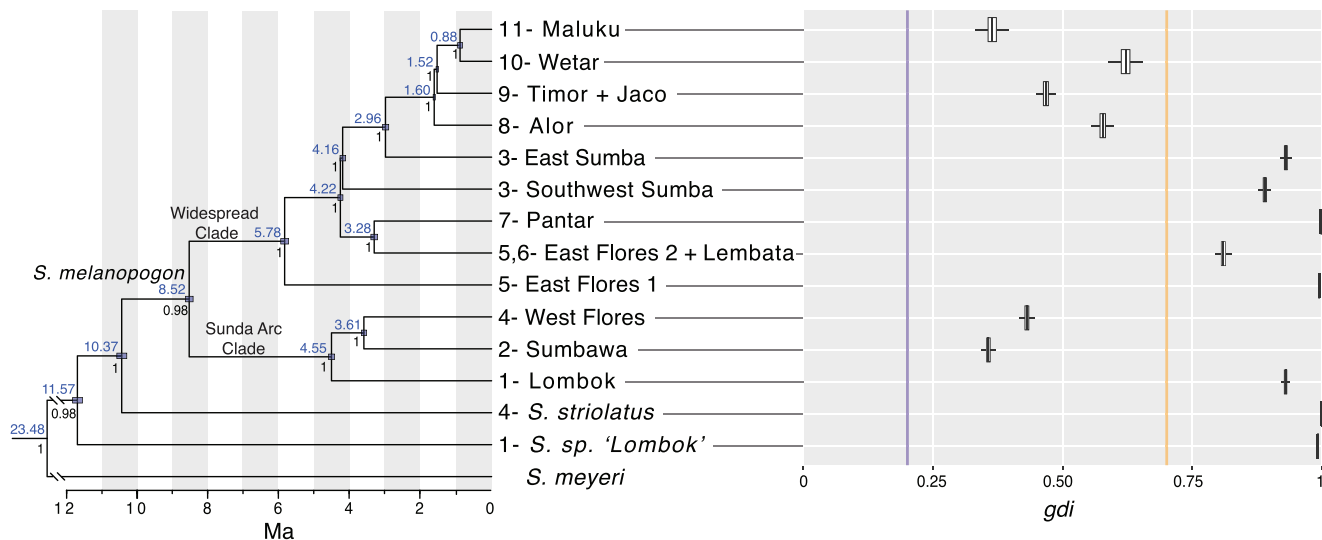
Figure 4. Species Tree and Genealogical Divergence Index. Time-calibrated multispecies coalescent phylogeny from BPP using all 1,154 loci and genealogical divergence index (*gdi*) values for each ingroup lineage. The numbers above the phylogenetic nodes represent mean node ages in millions of years (Ma), the numbers below nodes represent posterior probability support, and the node bars represent the 95% posterior density range for node age. Numbers before lineage names refer to the longitudinal ordering of islands shown in Figure 2. Box and whisker plots for *gdi* are calculated from theta and tau estimates from BPP, with values above 0.7 (marked with the orange line) indicating potential species status and values below 0.2 (marked with the purple line) indicating potential conspecific status. Intermediate values represent an ambiguous zone of species formation.

(mean = 0.51 million years ago) and between Lembata and East Flores 2 (mean = 0.26 million years ago).

PHYLOGENOMICS

Bootstrap support is high across the partitioned Maximum Likelihood phylogeny of the exon + flanking region dataset with BS = 100 for every major node (Fig. 3A). Most of the same clades described for the mitogenomic phylogeny are recovered, with a few exceptions. These include (1) a basal clade within *S. melanopogon* that includes Sumbawa and West Flores populations as sister lineages; (2) Pantar as the sister taxon of the East Flores 2 + Lembata lineage; (3) a monophyletic East Flores 2 clade; (4) Alor as sister to Timor + Wetar + Maluku (though with a very short branch length); and (5) the Maluku islands of Banda, Ai, and Kur as a monophyletic sister group of Wetar. The summary coalescent ASTRAL-III phylogeny produces the same relationships with local posterior probabilities of 1.0 for each node (Fig. S7A).

The time-calibrated BPP and STARBEAST2 Bayesian species tree analyses converged on the same topology as found in the ML and ASTRAL-III analyses (Fig. 4A; Fig. S7B). The BPP estimates of node ages are generally older than those from STARBEAST2, which may reflect the fact that we were able to include a substantially larger number of samples and sequence loci in the BPP analyses (1154 sequence loci for 101 individuals vs. 200 sequence loci for 36 individuals). Estimates for the divergence time of *Sphenomorphus* sp. "Lombok" and the rest of the Lesser Sunda

Sphenomorphus ranges from ~11.6 million years ago (BPP) to ~9.5 million years ago (STARBEAST2), followed by the divergence of *S. striolatus* and the *S. melanopogon* complex by ~10.4 million years ago (BPP) or ~8.2 million years ago (STARBEAST2). The initial split within *S. melanopogon* between the Lombok + Sumbawa + West Flores lineage and all other lineages is estimated to have occurred either ~8.5 million years ago (BPP) or ~6.2 million years ago (STARBEAST2). Some of the most recent divergence events based on the 200-locus dataset are estimated at ~0.18 million years ago between Lembata and East Flores 2 and ~0.36 million years ago between the Banda Islands and Kur.

The mean estimated *gdi* values ranged from ~0.36 to 0.99, with eight lineages receiving scores above 0.7 and six lineages in the ambiguous zone above 0.2 but less than 0.7 (Figure 4). The highest four *gdi* values, all >0.99, belong to *S.* sp. "Lombok", *S. striolatus*, East Flores 1, and Pantar. The ambiguous values belong to the sister lineages Lombok and West Flores and the clade containing the Alor, Timor+Jaco, Wetar, and Maluku lineages.

BIOGEOGRAPHY

Of the six biogeographic models, DIVALIKE + J generally provided the best fit to the data, followed by DEC + J (Table S6). The three best-supported models were obtained under the "PAIC" dispersal modifier, which takes into account historical and current connectivity between regions. The best joint history of the top supported DIVALIKE + J "PAIC" model (Fig. 3C;

Fig. S8) estimates an initial occupation of Lombok followed by dispersal eastward to West Flores. Subsequent dispersal from Flores westward to Sumbawa represents the divergence between *S. striolatus* and the common ancestor of *S. melanopogon*. Temporally, the next dispersal events within *S. melanopogon* occur from Sumbawa to East Flores, then from East Flores to Lembata, followed by nearly simultaneous dispersal from Sumbawa back to Lombok and from Lembata to Sumba. Remaining dispersal events within the Sunda Arc involve colonization of West Flores by Sumbawa and East Flores by Lembata. The Banda Arcs appear to have been colonized by two dispersal events, one from Lembata to Pantar and the other from Sumba to Timor, with subsequent dispersal events from Timor to Alor, from Timor to Wetar, and from Wetar to Maluku. The DEC + J "PAIC" model has the same best joint history as the top model, while the BAYARE-ALIKE + J "PAIC" model (Fig. S8) differs in that it estimates the first dispersal event to have occurred from Lombok to Sumbawa, with a subsequent dispersal from Sumbawa to West Flores (representing the divergence between *S. striolatus* and the common ancestor of *S. melanopogon*). The DIVALIKE + J "Distance" model is mostly consistent with the top model with a few exceptions: (1) Sumba is indicated to have been colonized by the East Flores 1 lineage, with subsequent dispersal from Sumba back to Lembata; and (2) the second dispersal event out of Sumba is to Alor (rather than to Timor) followed by dispersal from Alor to Timor (Fig. 3D).

POPULATION STRUCTURE

The first two components of the PCA for all Lesser Sunda and Maluku samples accounted for a combined 40.7% of the variance and recovered four distinct clusters corresponding to (1) *Sphenomorphus* sp. "Lombok", (2) *S. striolatus*, (3) a Sunda Arc clade of *S. melanopogon*, and (4) a widespread clade of *S. melanopogon* (Fig. 5A). The first two components of the PCA within the Sunda Arc clade captured 49.5% of the variance and recovered three tightly packed clusters corresponding to (1) Lombok, (2) Sumbawa, and (3) West Flores (Fig. 5B). The first two components of the PCA for the widespread clade explained 46.8% of the variance, with four major clusters of samples: (1) East Flores 1, (2) Southwest Sumba, (3) Pantar + East Flores 2 + Lembata, and (4) East Sumba + Timor + Jaco + Alor + Wetar + islands in Maluku. Pantar samples are slightly separated from East Flores 2 + Lembata, and East Sumba and Maluku are barely distinct from Timor + Jaco + Alor + Wetar (Fig. 5C). The first two components of the PCA for the eastern subclade accounted for 37.8% of the variance and separated the samples into three clusters: (1) Timor + Jaco, (2) Alor, and (3) Wetar + Maluku islands (Fig. S9).

Mantel tests for signatures of IBD were significant within *S. melanopogon* and the Sunda Arc Clade (simulated *p* values

<0.00001) and less so within the widespread clade (simulated *p* value = 0.00076) (Fig. S10). Visualization of the genetic vs. geographic distance plots shows distinct, patchy densities of points for *S. melanopogon* and the Widespread clade, indicating geographically distant and genetically divergent populations (Fig. 5D and F). The Sunda Arc plot also contains more than one cloud patch but is more consistent with a continuous cline than the other two sample groups (Fig. 5E).

STRUCTURE analyses for *S. melanopogon* returned an optimal K value of 2 (Table S5), roughly corresponding to the relationship of the Sunda Arc and widespread clades, with the East Flores 1 clade revealing a mixed ancestry between the two (Fig. 5G). The highest value of K to return meaningful clusters was 10, mostly concordant with the already defined clades from previous analyses, although within the eastern subclade, this method finds Alor distinct from Timor, Jaco, Wetar, and the Maluku islands. Within the Sunda Arc clade, the optimal K was 3 (Table S5), corresponding to Lombok, Sumbawa, and West Flores, although K = 5 revealed structure within Sumbawa and West Flores (Fig. 5H). Within the Widespread clade, the optimal K was 3, grouping (1) East Flores 1, SW Sumba, and E Sumba; (2) East Flores 2, Lembata, and Pantar; and (3) Alor, Timor, Jaco, Wetar, and the Maluku islands (Fig. 5I). Setting K = 7 also recovered meaningful clusters, although in contrast to K = 10 from the *S. melanopogon* analysis, it found Timor + Jaco to be distinct from the rest of the eastern subclade.

GENE FLOW AND INTROGRESSION

Population divergence times estimated with G-PHOCS were generally older than those estimated with BPP or STARBEAST2 (though are more similar to the BPP results), ranging from ∼1.8 Ma for Wetar and the Maluku Islands, up to ∼12.1 million years ago for *Sphenomorphus* sp. "Lombok" and Lombok (Fig. 6). Estimates of migrants per generation were effectively zero for most lineage comparisons. Three comparisons recovered nonzero migration values, including (1) between East Flores 2 and Lembata, (2) between Wetar and the Maluku Islands, and (3) between West Flores and Sumbawa. All migration bands exhibit fewer than 0.5 migrants per generation, a threshold that we and others have previously suggested as a species threshold (Shaffer and Thompson, 2007; Reilly et al., 2021). Asymmetric migration is found predominantly from East Flores 2 into Lembata and from Wetar into Maluku (Fig. 6).

Four-taxon tests returned significant signals of ancient introgression between the West Flores clade and the East Flores 2 clade ($Z = 10.08$, $p <0.00001$), between *S. striolatus* and the West Flores clade ($Z = −6.95$, $p <0.00001$), and between the Lombok and Sumbawa clades ($Z = 6.08$, $p <0.0001$) (Table 1; Fig. S11A–I). A near-significant result was estimated between the Alor and Pantar clades. None of the five-taxon tests returned
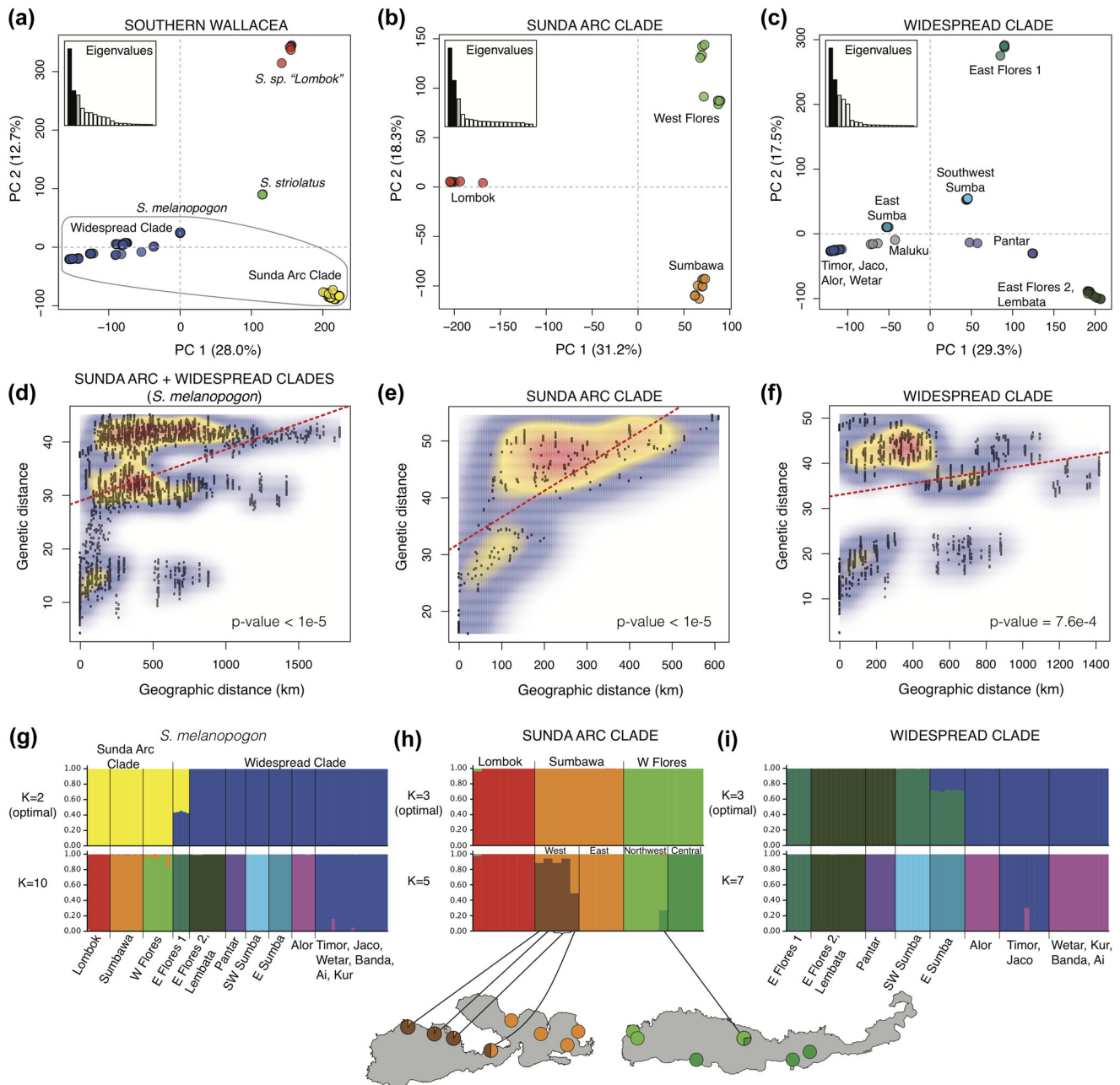
Figure 5. Principal component analysis (PCA) of 171,751 SNPs for (A) the southern Wallacean *Sphenomorphus* assemblage, (B) the Sunda Arc clade of *S. melanopogon*, and (C) the Widespread clade of *S. melanopogon*. Scatterplots depict isolation-by-distance results for (D) *S. melanopogon* and its two major intraspecific groups, (E) the Sunda Arc Clade and (F) the Widespread Clade, plotted as geographic versus genetic distance, with each point representing a two-sample comparison. Point density is measured with a 2-dimensional kernel density estimation, where blue colors represent lower densities and red colors indicate the highest densities. The lower panels depict STRUCTURE genetic clustering results of the optimal K and the largest K that returned biogeographically meaningful results for (G) *S. melanopogon*, (H) the Sunda Arc clade, and (I) the Widespread clade.

significant signals of introgression, although near-significant results were recovered for introgression between the Pantar clade and the ancestor of the Timor and Alor clades and between the West Flores clade and the Southwest Sumba clade (Table 2; Fig. S11J–L).

Discussion

This study has uncovered a cryptic radiation of *Sphenomorphus* forest skinks in the oceanic islands of southern Wallacea. Our estimates of population structure, gene flow, and genealogical divergence provide multiple lines of evidence supporting the
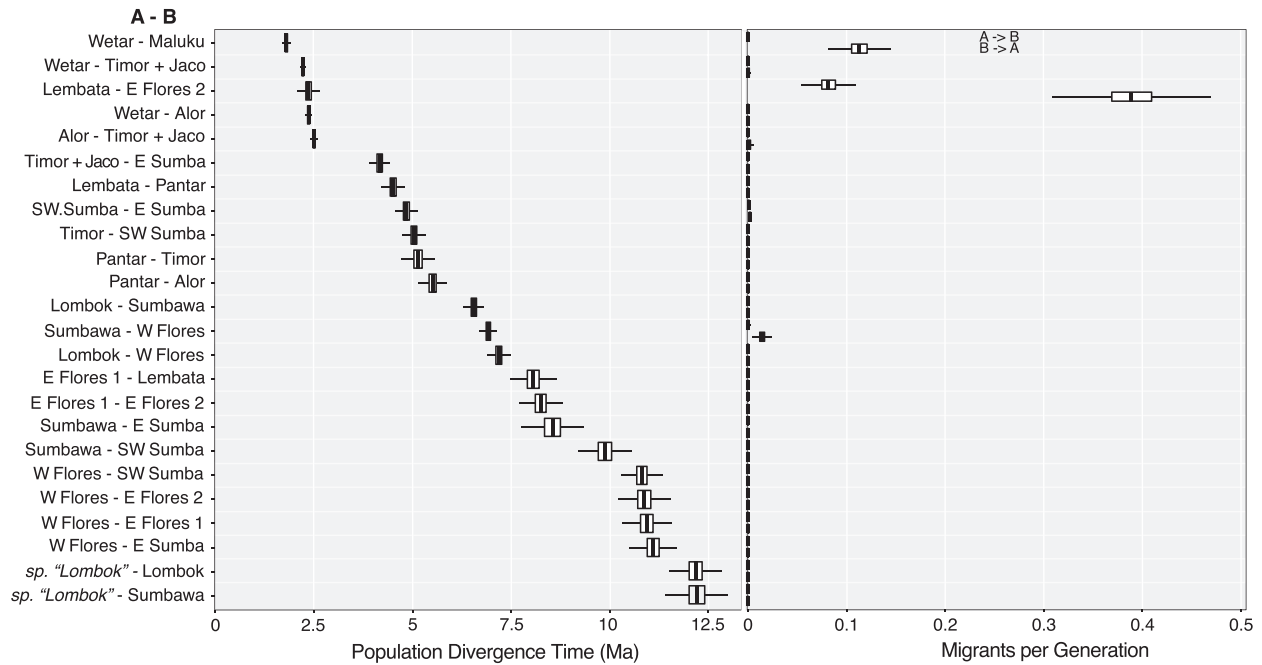
Figure 6. Box and whisker plots of the posterior probability distribution of population divergence time (millions of years ago) and migrants per generation from G-PHOCS for adjacent or overlapping regional lineages.

Table 1. Four-taxon tests for introgression with significant results and lineages in bold. N = number of sample comparisons, ABBA/BABA = number of allelic configurations ± standard deviation, D = Patterson's D-statistic, Z = statistical confidence score calculated from the mean and standard deviation of the individual D-statistics, P = p value. *S.me* = *S. meyeri*, Lom1 = *S. sp. "Lombok"*, *S.st* = *S. striolatus*, Lom2 = Lombok, Sw = Sumbawa, WFl = West Flores, EFl1 = East Flores 1, EFl2 = East Flores 2, Le = Lembata, Ti = Timor, We = Wetar, Al = Alor, Pa = Pantar, SWSu = Southwest Sumba, ESu = East Sumba, Ma = Maluku.

N	(((P1, P2), P3), O)	ABBA	BABA	D	Z	P
450	(((EFl2, EFl1), WFl), *S.st*)	835.6 ± 40.2	1127.9 ± 56.8	−0.148 ± 0.015	**10.08**	**<10E-10**
630	(((EFl1, WFl), *S.st*), Lom1)	923.0 ± 51.2	767.0 ± 36.6	0.092 ± 0.013	**−6.95**	**<10E-10**
1260	(((WFl, Sumbawa), Lom2), *S.st*)	1232.2 ± 120.5	832.6 ± 64.8	0.192 ± 0.032	**−6.08**	**1.20E-09**
2352	(((Ti, We), Al), ESu)	246.2 ± 27.1	182.5 ± 20.5	0.148 ± 0.064	−2.33	0.02
98	(((Lom2, *S.st*), Lom1), *S.me*)	1377.5 ± 62.4	1288.1 ± 53.9	0.033 ± 0.025	−1.32	0.19
1050	(((Le, EFl2), EFl1), Lom2)	180.1 ± 13.4	160.4 ± 14.0	0.058 ± 0.054	−1.08	0.28
900	(((EFl2, Le), Pa), EFl1)	169.3 ± 60.5	183.6 ± 61.2	−0.047 ± 0.065	0.72	0.47
2058	(((Al, ESu), SWSu), Pa)	505.4 ± 188.9	487.9 ± 164.1	0.007 ± 0.051	−0.14	0.89
2016	(((Ma, We), Ti), Al)	70.6 ± 23.1	73.4 ± 32.7	−0.001 ± 0.140	<0.01	0.99

Table 2. Five-taxon tests for introgression with significant results (P <0.001) highlighted in bold. N = number of sample comparisons, Z = statistical confidence score derived from the mean and standard deviation of individual D-statistics, P = p value. Lineage abbreviations are found in the Table 1 caption.

N	(((P1, P2), (P3, P4)), O)	Z (DFO)	P (DFO)	Z (DIL)	P (DIL)	Z (DFI)	P (DFI)	Z (DOL)	P (DOL)
1875	(((Ti, Al), (Pa, Le)), EFl1)	−2.94	0.003	−2.99	0.003	−0.06	0.948	0.07	0.945
1250	(((EFl2, EFl1), (WFl, Lo2)), *S.st*)	−1.69	0.090	−0.63	0.528	0.40	0.690	1.32	0.188
2592	(((Esu, SWSu), (WFl, Sum)), *S.st*)	−3.11	0.002	−3.17	0.002	2.83	0.005	3.03	0.002

recognition of these geographically structured lineages as species. Because there is broad agreement that the tropics harbor high biodiversity and that oceanic archipelagos promote high levels of endemism (e.g., Kier et al. 2009), it is not surprising that our examination of a wide-ranging species complex across a tropical oceanic archipelago revealed genetically distinct lineages, most of which are single-island endemics. What is surprising is the sheer number of such lineages given the overall lack of morphological variation among these populations, the unexpected geographical locations of lineage boundaries, and the phylogenetic relationships inferred for these cryptic lineages. We reveal a complex biogeographical history that includes evidence of both short- and long-distance overwater dispersal events, crossings of temporary land bridges, and secondary contact of lineages following the merger of paleo-islands. Analyses of gene flow, introgression, and genealogical divergence levels indicate that the *Sphenomorphus melanopogon* complex is composed of at least nine species and likely many more. Below, we discuss the issue of species delimitation as it pertains to island archipelagos, as well as the biogeographical, conservation, and taxonomic implications of our findings.

SPECIES DELIMITATION IN AN ISLAND ARCHIPELAGO

Given the discontinuous nature of land and sea in island archipelagos, it is unsurprising that allopatric divergence is a key process driving speciation in these settings. It is also true that allopatric lineages serve as the primary points of contention for the so-called "species problem" (Padial and De la Riva 2021). For some researchers (typically proponents of the biological species concept; Mayr 1942; Butlin and Stankowski 2020), the evolution of intrinsic reproductive isolating mechanisms is the essential feature of speciation. In our view (as proponents of the Evolutionary or General Lineage Concept of Species; e.g., Wiley 1978; De Queiroz, 1998, 2007), species are well-defined independent lineages on their own evolutionary trajectories. Such lineages can exist with or without intrinsic reproductive isolating mechanisms, although we agree that reproductive incompatibility is the most critical feature of the temporally extended process of speciation in that it provides fixation (de Queiroz 2007). The presence of intrinsic reproductive isolating mechanisms and lineage independence is typically self-evident for sympatric lineages, but for allopatric populations, we are generally left to make a judgment call about morphological divergence as a proxy for either reproductive incompatibility (in the case of "biological species") or lineage isolation (in the case of "evolutionary species").

The genomics revolution makes possible an alternative approach to species delimitation for advocates of lineage-based concepts in that the presence or absence and extent of gene flow can be quantified directly, reducing our reliance on morphologi-

cal proxies in the assessment of lineage status (Fujita et al. 2012; Burbrink and Ruane 2021; Leaché et al. 2021). Our approach emphasizes barriers to gene flow by estimating rates of migration across putative lineage boundaries and treating those for which population demographic analyses infer fewer than 0.5 migrants per generation as distinct lineages. Because there can be "gray zones" in the speciation process that render species delimitation both challenging and somewhat subjective (Roux et al. 2016), we emphasize that we do not consider 0.5 migrants per generation as a fixed threshold for actual species status. Instead, we recognize it as a suitable indicator of lineage independence because it is the threshold migration rate above which gene flow is expected to result in genomic homogenization under a model of genetic drift (Wright 1931). It is important to recognize that this approach is very much dependent on sufficient sampling, particularly around contact zones, and could be confounded by sampling gaps (some of which characterize our dataset, such as for Central Flores), particularly if there is a pattern of isolation by distance (see Hillis et al. 2021). The criterion of 0.5 migrants per generation as an indicator of lineage status was suggested by Shaffer and Thompson (2007) and has been previously put into practice by Barley et al. (2013) and Reilly et al. (2022a). While we recognize that our approach may not satisfy those focused on the evolution of intrinsic reproductive isolating mechanisms as the underlying rationale for recognizing species, we note that it is simply impractical to empirically test for reproductive isolation among large numbers of genetically divergent allopatric populations. Our approach is practical, tolerant of different species concepts, amenable to studies of cryptic diversification, and provides the opportunity for producing well-reasoned taxonomic decisions.

Employing this approach and taking into account instances of sympatry between lineages (Table 3), we present evidence for as many as 11 *Sphenonomorphus* species in Lesser Sundas, including nine cryptic lineages within *S. melanopogon*. Reilly et al. (2019a, 2022a) also found elevated numbers of cryptic species in Lesser Sunda fanged frogs (genus *Limnonectes*) and flying lizards (genus *Draco*). In those two genera, we did not find evidence of sympatry among cryptic lineages, and we thus could not assess the presence of intrinsic barriers to gene flow that would satisfy the biological species concept. However, in *Sphenomorphus*, we have identified sympatric cryptic species within *S. melanopogon* (the East Flores 1 and East Flores 2 lineages; Table 3), suggesting that a full application of the BSC would support multiple species within this assemblage. This begs the question of just how extensive intrinsic reproductive incompatibilities may be across this species complex.

TAXONOMIC IMPLICATIONS

Our analyses confirm that the southern Wallacean *Sphenomorphus* assemblage includes *Sphenomorphus* sp. "Lombok,"

Table 3. Summary of sympatric or suspected sympatric lineages within the southern Wallacean *Sphenomorphus* group. Locality # refers to the localities in Figure S1 where the lineages were collected sympatrically on our expeditions. MtDNA divergence estimates represent uncorrected sequence divergence for the mitochondrial genome listed in Table S4. Nuclear DNA divergence estimates are taken from the BPP tree in Figure 4. Introgression refers to significant four- or five-taxon tests, and Gene Flow refers to any detectable migration from G-PHOCS analyses.

Lineage 1	Lineage 2	Sympatric	Localitiy #	mtDNA div. (%)	nDNA div. (Ma)	Introgression/Gene Flow
S. sp. "Lombok"	Lombok	Yes	3	18.0%	11.6	No/No
S. striolatus	West Flores	Yes	22	17.9%	10.4	Yes/No
East Flores 1	East Flores 2 - Lembata	Yes	26, 27	12.9%	5.8	No/No
Southwest Sumba	East Sumba	Maybe	-	11.8%	4.2	No/No

S. striolatus, and the *S. melanopogon* complex, the latter having proven to be a challenge for morphology-based species delimitation (Shea 2012). Within *melanopogon*, we interpret our suite of results as supporting nine species (Table 4), which are discussed here. Within the *S. melanopogon* Sunda Arc clade, three candidate species from Lombok, Sumbawa, and West Flores are supported by genomic and mitochondrial monophyly (Fig. 3), deep divergence (>2 million years ago; Figs. 4 and 6), unambiguous separation according to genomic PCA and population structure plots (Fig. 5B, G, and H), and $2Nm$ (migrants per generation) estimates close to zero (<0.01) between adjacent lineages (Fig. 6). Although four-taxon introgression tests between the Lombok and Sumbawa lineages indicated that ancient gene flow occurred (Table 1), and *gdi* values (Fig. 4) suggest that Sumbawa and West Flores are in the ambiguous speciation zone, we find that the Sunda Arc clade contains three species.

The Widespread clade includes six candidate species, namely, the East Flores 1, Pantar, East Flores 2 + Lembata, Southwest Sumba, East Sumba, and Eastern lineages. All six are supported by genomic and mitochondrial monophyly (Fig. 3), deep divergence times (Figs. 4, 6), separation within PCA and population structure plots (Figs. 5C, G, and I), and $2Nm$ estimates of nearly zero (Fig. 6). Our conclusion that these lineages represent species is bolstered by the fact that the two East Flores lineages occur in direct syntopy (Table 3) and the lack of gene flow between the two Sumba lineages despite their close geographical proximity and apparent lack of dispersal barriers between them. Even the closely related East Flores 2 and Lembata sublineages have low gene flow ($2Nm$ <0.5 in both directions) but a shorter divergence time (~0.18 million years ago), and continuing physical connectivity between the two regions suggests that we conservatively treat them as conspecifics.

Within the Eastern clade, a similar confounding situation exists, with four subclades (Alor, Timor + Jaco, Wetar, and Maluku) initially considered candidate species. Their distinctive-

ness is supported by genomic monophyly (Fig. 3) and $2Nm$ values <0.5 (Fig. 6) and mitochondrial monophyly for three of the groups (but not for Maluku). However, the divergence times are short relative to the other candidate species in the archipelago (<1 Ma), the PCA suggests three groups (Fig. S9), population structure plots suggest two groups (Fig. 5I), and *gdi* values are ambiguous (Fig. 4). Given the lack of agreement across analyses, we suggest that the Eastern clade be considered a single, highly structured species, pending more comprehensive analysis.

Populations of the *S. melanopogon* complex that we were unable to include in this study occur on the islands of Adonara, Babar, Damar, Komodo, Rinca, Rote, Sumba (West Sumba), and Timor (West Timor) and the islands of Tinjil and Deli off of southwest Java, whose colonization by this group of skinks is rather mysterious (Shea 2012). These areas, along with smaller, lesser-known islets, almost certainly hold even more species-level diversity than described in this study. Continued surveys and a re-evaluation of morphological variation, guided by the results of this study, will be needed to further refine species boundaries within Wallacean *Sphenomorphus*. Our published mitogenomic dataset provides a useful resource for the identification of newly acquired samples, as it is now possible to apply a DNA barcoding approach to these samples using any coding mitochondrial gene.

GEODYNAMICS AS A DRIVER OF DIVERSIFICATION

The term "species pump" was first used to explain how cyclical forest fragmentation and connection driven by Pleistocene climate cycles were instrumental in creating the high species diversity of Amazonian birds (Haffer 1969). The idea of fluctuating connectivity and vicariance as a driver of speciation has also been applied to island systems that repeatedly merge and fragment due to Pleistocene sea-level changes or to so-called "Pleistocene Aggregate Island Complexes" (PAICs; Inger 1954; Heaney 1985; Brown and Diesmos 2002; Brown and Diesmos

Table 4. Combined analyses considering candidate species within *S. melanopogon*. The mtDNA species delim column refers to candidate species identified by ABGD in the mitochondrial genome dataset, where Y = yes the data support the distinctiveness of this clade, N = no the data do not support the distinctiveness of this clade, and numbers in parentheses next to Y indicate the number of candidate species identified within that clade.

S. melanopogon candidate species	mtDNA monophyly	mtDNA species delim	nDNA monophyly	gdi	nDNA Div Time (Ma)	PCA cluster	STRUCTURE	Gene Flow into pop	Final sp. inference
Lombok	Y	Y	Y	Y	4.6	Y	Y	Absent	1
Sumbawa	Y	Y	Y	Ambiguous	3.6	Y	Y	Absent	2
West Flores	Y	Y (3)	Y	Ambiguous	3.6	Y	Y	Absent	3
East Flores 1	Y	Y	Y	Y	5.8	Y	Y	Absent	4
Pantar	Y	Y (3)	Y	Y	3.3	Y	Y	Absent	5
East Flores 2	N	N	N	N	2.5	N	N	Low	6
Lembata	Y	N	Y	N	2.5	N	N	Moderate	6
Southwest Sumba	Y	Y	Y	Y	4.2	Y	Y	Absent	7
East Sumba	Y	Y	Y	Y	4.2	Y	Y	Absent	8
Alor	Y	Y	Y	Ambiguous	1.6	Y	N	Absent	9
Timor + Jaco	Y	Y (2)	Y	Ambiguous	1.5	Y	Y	Absent	9
Wetar	Y	N	Y	Ambiguous	0.9	N	N	Low	9
Banda + Ai	Y	Y	N	N	0.4	N	N	Absent	9
Kur	Y	N	Y	N	0.4	N	N	Absent	9

2009). The hypothesis of PAIC-driven speciation has been supported in some cases (Heaney 1986), determined to be only partially responsible in some systems (Esselstyn and Brown 2009; Esselstyn et al. 2009; Brown et al. 2013; Papadopoulou and Knowles 2015), and inferred to not be a factor in others (Bringloe and Saunders 2018; Oaks et al. 2019; Ebdon et al. 2021; Jiang et al. 2021). Some studies have even concluded that the oscillating connectivity of islands will result not in a species pump but in a species vacuum, in which signatures of incipient lineage divergence are periodically erased by lineage mergers (Papadopoulou and Knowles 2017; Jiang et al. 2021). However, the relative recency of sea level fluctuations within PAIC systems may be less relevant for species formation than intermediate sea levels associated with more long-lasting archipelago configurations, such that species divergences are frequent *between* PAICs and less so *within* them (Norder et al. 2019).

While our biogeographical model testing favored a model incorporating increased dispersal between PAIC islands, our results do not support a PAIC species pump process within Wallacean *Sphenomorphus* since the timing of most divergences well predates the Pleistocene and most of the shallower divergence times (such as in the Eastern Clade) occur on islands that do not become connected to one another during glacial cycles. Two examples illustrate the extremes. Within the "Greater Flores" PAIC, the East Flores 2 and Lembata lineages are closely related and have experienced moderate gene flow, illustrating the potential influence of temporary land bridges in preventing (rather than promoting) divergence. At the other extreme, the lack of gene flow and deep divergence between discrete *S. melanopogon* lineages on Lombok and Sumbawa (which together compose the Lombok + Sumbawa PAIC) suggest that species formation on these islands was complete prior to the start of the Pleistocene. Close examination of other Wallacean archipelagos, such as Sulawesi and the Philippines, suggests that precursor paleo-islands allowed early colonizing lineages to greatly diverge in isolation in a landscape barely resembling that of the present day (Evans et al. 2003; Jansa et al. 2006; Blackburn et al. 2010; Setiadi et al. 2011; Stelbrink et al. 2012; Brown and Siler 2014). This may apply to the Lesser Sundas as well, as continuous subduction-driven volcanism and uplift over the past 10+ million years in the Sunda Arc portion of the archipelago suggests that the current island configuration may not accurately reflect the distribution of land and sea 10, 5, or even 2 million years ago. Thus, the contemporary distance between islands and/or their inferred connectivity based on channel depths may have little explanatory power regarding diversification patterns of early colonizing lineages that predate the Pleistocene, as seen in other lizard radiations in the region (Barley et al. 2013; Tallowin et al. 2018; Oaks et al. 2019).

CONSERVATION OF WALLACEAN ENDEMISM HOTSPOTS

The natural habitats of the Indo-Australian Archipelago, including those of Wallacea, are among the most threatened in the world (Bickford et al., 2007; Lohman et al. 2011). These habitats buffered many terrestrial taxa against extreme environmental perturbations during glacial periods, contributing to insular species formation and maintenance by reducing the extinction of geographically isolated populations (Fordham and Brook 2010). However, these historically stable and ecologically reliable natural habitats and their species are highly vulnerable to nonnatural human impacts. Because terrestrial vertebrate island endemism is over eight times greater than mainland endemism, ongoing habitat loss on islands promotes the extinction of island endemics, which already comprise the majority of human-caused extinctions (Kier et al. 2009; Fordham and Brook 2010; Whittaker et al. 2017). Just to the southwest of Lesser Sundas lies Christmas Island, where both recognized and unrecognized species or insular lineages have either gone extinct or are now threatened, highlighting the vulnerability of the region's insular herpetofauna (Smith et al., 2012; Oliver et al. 2018)

Biodiversity and endemism patterns, commonly used to assess conservation priorities and action plans, are dependent upon the taxonomic recognition of relevant groups and the spatial scale of biodiversity sampling (Daru et al. 2020). Recent fine-scale studies of widespread lineages within Lesser Sundas have revealed endemism hotspots that were not previously known. For example, eastern Flores, which may have existed as a separate island in the past, is a previously unrecognized endemism hotspot (Reilly et al. 2019a; Reilly et al. 2022a; this study). Future work should integrate data from across as many taxa as possible across the Lesser Sundas to identify and clarify areas of endemism and shared evolutionary history (Rosauer et al. 2009; Leaché et al. 2020), and these areas should be prioritized when new protected areas are planned. Additionally, many of the smaller unsampled islands in southern Wallacea surely have undocumented or undescribed small-range endemic species, and knowledge of their existence could be used to bolster the argument for designating new protected areas. In the case of *Sphenomorphus*, whether names are given to each distinct lineage, they are all clearly evolutionarily significant units (ESUs; Moritz 1994) that should be conserved.

Conclusions

This study has revealed the existence of multiple deeply divergent lineages of *Sphenomorphus* skinks in the Lesser Sunda Archipelago, including single-island endemics, lineages that span multiple islands, and lineages apparently restricted to segments of individual islands. The most likely colonization scenarios indicate a steady stream of dispersal events throughout the archipelago over the past 10–12 million years, with some lineages having come back into secondary contact after long periods of isolation. In some cases, these lineages appear to form narrow contact zones, while in others, the lineages have become syntopic, confirming pre- or postzygotic reproductive isolation and suggesting that other similarly divergent lineages may also be both reproductively and ecologically distinct. The assemblage is composed of as many as 11 species (and perhaps more), most of which have relatively limited distributions, but with one wide-ranging species occurring on most of the Banda Arc islands and having a distribution spanning nearly 1000 km. Our results suggest that the islands of Lesser Sundas may have not only been arranged differently in the past but may also have more heterogeneous temporal histories than currently understood. This applies most notably to Sumbawa, Flores, and Sumba, all of which have biogeographical signatures suggesting that they were formed via the merger of distinct palaeo-islands. Southern Wallacea has a diverse and endemic fauna, much of which is not yet recognized taxonomically. Phylogenomic and population genomic analyses of Lesser Sundas species complexes, such as *Sphenomorphus*, identify not only new species but also regional endemism hotspots that lack conservation protection.

CONFLICT OF INTEREST
The authors declare no conflict of interest.

DATA ARCHIVING

Sequence data for the ND4 mitochondrial gene have been submitted to the GenBank database under accession numbers MW291696–291940. Sequencing data from the exon-capture experiment have been submitted to the NCBI sequence read archive under accession number PRJNA682772. Software input files are available from the Dryad Digital Repository at https://doi.org/10.7291/D1XX09.

LITERATURE CITED

Alcala, A.C. (1966). Populations of three tropical lizards on Negros Island, Philippines. Stanford University.

Ali, J.R. & Heaney, L.R. (2021) Wallace's line, Wallacea, and associated divides and areas: history of a tortuous tangle of ideas and labels. *Biol. Rev.*, 96, 922–942.

Allio, R., Donega, S., Galtier, N. & Nabholz, B. (2017) Large variation in the ratio of mitochondrial to nuclear mutation rate across animals: implications for genetic diversity and the use of mitochondrial DNA as a molecular marker. *Molecular Biology and Evolution*, 34, 2762–2772.

Arèvalo, E., Davis, S.K. & Sites, J.W., Jr. (1994) Mitochondrial DNA sequence divergence and phylogenetic relationships among eight chromosome races of the *Sceloporus grammicus* complex (Phrynosomatidae) in central Mexico. *Systematic Biology*, 43, 387–418.

Auffenberg, W. (1980) The herpetofauna of Komodo, with notes on adjacent areas. *Bull. Florida State Mus., Biol. Sci.*, 25, 39–156.

Barley, A.J., White, J., Diesmos, A.C. & Brown, R.M. (2013) The challenge of species delimitation at the extremes: diversification without morphological change in Philippine sun skinks. *Evolution*, 67, 3556–3572.

Bi, K., Vanderpool, D., Singhal, S., Linderoth, T., Moritz, C. & Good, J.M. (2012) Transcriptome-based exon capture enables highly cost-effective comparative genomic data collection at moderate evolutionary scales. *BMC Genomics*, 13, 403.

Bickford, D.P., Supriatna, J., andayani, N., Iskandar, D.T., Evans, B.J., Brown, R.M. et al. (2007) Indonesia's protected areas need more protection – suggestions from island examples. Pgs 53–78 In Sodhi, N. S., G. Acciaioli, M. Erb, and A. K.-J. Tan (Eds.) Biodiversity and human livelihoods in protected areas. Cambridge Press, Cambridge.

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, 10.

Blom, M.P., Horner, P. & Moritz, C. (2016) Convergence across a continent: adaptive diversification in a recent radiation of Australian lizards. *Proc. R. Soc. B Biol. Sci.*, 283, 20160181.

Blom, M.P., Matzke, N.J., Bragg, J.G., Arida, E., Austin, C.C., Backlin, A.R. et al. (2019) Habitat preference modulates trans-oceanic dispersal in a terrestrial vertebrate. *Proc. R. Soc. B Biol. Sci.*, 286, 20182575.

Brandley, M.C., Wang, Y., Guo, X., Nieto Montes de Oca, A., Feria-Ortiz, M. et al. (2011) Accommodating heterogenous rates of evolution in molecular divergence dating methods: an example using intercontinental dispersal of *Plestiodon* (*Eumeces*) lizards. *Systematic Biology*, 60, 3–15.

Bringloe, T.T. & Saunders, G.W. (2018) Trans-Arctic speciation of Florideophyceae (Rhodophyta) since the opening of the Bering Strait, with consideration of the "species pump" hypothesis. *Journal of Biogeography*, 46, 694–705.

Brown, R.M. & Diesmos, A.C. (2002) Application of lineage-based species concept to oceanic island frog populations: the effects of differing taxonomic philosophies on the estimation of Philippine biodiversity. *Silliman J.*, 42, 133–162.

———. (2009) Philippines biology. Pp. 723–732, in R. Gillespie and D. Clague, eds. Encyclopdia of Islands. Berkeley, California, USA: University of California Press.

Brown, R.M. & Siler, C.D. (2014) Spotted stream frog diversification at the Australasian faunal zone interface, mainland versus island comparisons, and a test of the Philippine 'dual-umbilicus' hypothesis. *Journal of Biogeography*, 41, 182–195.

Brown, R.M., Siler, C.D., Oliveros, C.H., Esselstyn, J.A., Diesmos, A.C., Hosner, P.A. et al. (2013) Evolutionary processes of diversification in a model island archipelago. *Ann. Rev. Ecol. Evol. Syst.*, 44, 411–435.

Brown, W.C. (1991) Lizards of the genus *Emoia* (Scincidae) with observations on their evolution and biogeography. *Mem. Calif. Acad. Sci.*, 15, 1-94.

Burbrink, F.T. & Ruane, S. (2021) Contemporary philosophy and methods for studying speciation and delimiting species. *Ichthyol. Herpetol.*, 109, 874–894.

Butlin, R.K. & Stankowski, S. (2020) Is it time to abandon the biological species concept? *No. Natl. Sci. Rev.*, 7, 1400–1401.

Chernomor, O., von Haeseler, A. & Minh, B.Q. (2016) Terrace aware data structure for phylogenomic inference from supermatrices. *Systematic Biology*, 65, 997–1008.

Darlington, P.J. (1957) Zoogeography: the Geographical Distribution of Animals. Cambridge, Massachusetts, USA: Harvard University Press.

Daru, B.H., Farooq, H., Antonelli, A. & Faurby, S. (2020) Endemism patterns are scale dependent. *Nature Communication*, 11, 2115.

De Queiroz, K. (1998) The general lineage concept of species, species criteria, and the process of speciation. Pp. 57–75 in D. J. Howard and S. H. Berlocher eds. Endless forms: Species and speciation. Oxford Univ. Press, Oxford, United Kingdom.

———. (2007) Species concepts and species delimitation. *Systematic Biology*, 56, 879–886.

Deng, W., Maust, B.S., Nickle, D.C., Learn, G.H., Liu, Y., Heath, L. et al. (2010) DIVEIN: a web server to analyze phylogenies, sequence divergence, diversity, and informative sites. *Biotechniques*, 48, 405–408.

Duméril, A.M.C. & Bibron, G. (1839) Erpétologie Générale ou Histoire Naturelle Complète des Reptiles. Tome Cinquième. Contenant l'Histoire de Quatre-vingt-trois Genres et de Deux Cent Sept Espèces des Trois Dernières Familles de l'Ordre des Sauriens, Savoir: les Lacertides, les Chalcidiens et les Scincoïdiens. Paris, France, Libraire Encyclopédique de Roret.

Durand, E.Y., Patterson, N., Reich, D. & Slatkin, M. (2011) Testing for ancient admixture between closely related populations. *Molecular Biology and Evolution*, 28, 2239–2252.

Earl, D.A. & vonHoldt, B.M. (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Res.*, 4, 359–361.

Eaton, D.A. & Ree, R.H. (2013) Inferring phylogeny and introgression using RADseq data: an example from flowering plants (Pedicularis: Orobanchaceae). *Systematic Biology*, 62, 689–706.

Eaton, D.A., Hipp, A.L., González-Rodríguez, A. & Cavender-Bares, J. (2015) Historical introgression among the American live oaks and the comparative nature of tests for introgression. *Evolution*, 69, 2587–2601.

Ebdon, S., Laetsch, D.R., Dapporto, L., Hayward, A., Ritchie, M.G., Dincă, V. et al. (2021) The Pleistocene species pump past its prime: evidence from European butterfly sister species. *Molecular Ecology*, 30, 3575–3589.

Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.*, 32, 1792–1797.

Esselstyn, J.A. & Brown, R.M. (2009) The role of repeated sea-level fluctuations in the generation of shrew (Soricidae: *Crocidura*) diversity in the Philippine Archipelago. *Molecular Phylogenetics and Evolution*, 53, 171–181.

Esselstyn, J.A., Timm, R.M. & Brown, R.M. (2009) Do geological or climatic processes drive speciation in dynamic archipelagos? The tempo

and mode of diversification in Southeast Asian shrews. *Evolution*, 63, 2595–2610.

Esselstyn, J.A., Oliveros, C.H., Moyle, R.G., Peterson, A.T., McGuire, J.A. & Brown, R.M. (2010) Integrating phylogenetic and taxonomic evidence illuminates complex biogeographic patterns along Huxley's modification of Wallace's Line. *Journal of Biogeography*, 37, 2054–2066.

Evanno, G., Regnaut, S. & Goudet, J. (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, 14, 2611–2620.

Evans, B.J., Brown, R.M., McGuire, J.A., Supriatna, J., Andayani, N., Diesmos, A. et al. (2003) Phylogenetics of fanged frogs: testing biogeographical hypotheses at the interface of the Asian and Australian faunal zones. *Systematic Biology*, 52, 794–819.

Fordham, D.A. & Brook, B.W. (2010) Why tropical island endemics are acutely susceptible to global change. *Biodiversity and Conservation*, 19, 329–342.

Fujita, M.K., Leaché, A.D., Burbrink, F.T., McGuire, J.A. & Moritz, C. (2012) Coalescent-based species delimitation in an integrative taxonomy. *Trends in Ecology & Evolution*, 27, 480–488.

Green, R.E., Krause, J., Briggs, A.W., Maricic, T., Stenzel, U., Kircher, M. & Pääbo, S. (2010) A draft sequence of the Neandertal genome. *Science*, 328, 710–722.

Grismer, L.L., Wood, P.L., Jr., Poyarkov, N.A., Le, M.D., Kraus, F., Agarwal, I. et al. (2021) Phylogenetic partitioning of the third-largest vertebrate genus in the world, *Cyrtodactylus* Gray, 1827 (Reptilia; Squamata; Gekkonidae) and its relevance to taxonomy and conservation. *Vert. Zool.*, 71, 101–154.

Gronau, I., Hubisz, M.J., Gulko, B., Danko, C.G. & Siepel, A. (2011) Bayesian inference of ancient human demography from individual genome sequences. *Nature Genetics*, 43, 1031–1034.

Gwee, C.Y., Christidis, L., Eaton, J.A., Norman, J.A., Trainor, C.R., Verbelen, P. et al. (2017) Bioacoustic and multilocus DNA data of *Ninox* owls support high incidence of extinction and recolonization on small, low-lying islands across Wallacea. *Molecular Phylogenetics and Evolution*, 109, 246–258.

Haffer, J. (1969) Speciation in Amazonian forest birds. *Science*, 165, 131–137.

Hall, R. (2009) Southeast Asia's changing paleogeography. *Blumea*, 54, 148–161.

———. (2011) Australia-SE Asia collision: plate tectonics and crustal flow. *Geol. Soc. London Spec. Publ.*, 355, 75–109.

Heaney, L.R. (1985) Zoogeographic evidence for middle and late Pleistocene land bridges to the Philippine islands. *Mod. Quatern. Res. Southeast Asia*, 9, 127–144.

———. (1986) Biogeography of mammals in SE Asia: estimates of rates of colonization, extinction and speciation. *Biological Journal of the Linnean Society*, 28, 127–165.

Hillis, D.M., Chambers, E.A. & Devitt, T.J. (2021) Contemporary methods and evidence for species delimitation. *Ichthyol. Herpetol.*, 109, 895–903.

Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q. & Vinh, L.S. (2018) UFBoot2: Improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution*, 35, 518–522.

Inger, R.F. (1954) Systematics and zoogeography of Philippine Amphibia. *Fieldiana Zool.*, 33, 182–531.

Jackson, N.D., Carstens, B.C., Morales, A.E. & O'Meara, B.C. (2017) Species delimitation with gene flow. *Systematic Biology*, 66, 799–812.

Jansa, S.A., Barker, F.K. & Heaney, L.R. (2006) The pattern and timing of diversification of Philippine endemic rodents: evidence from mitochondrial and nuclear gene sequences. *Systematic Biology*, 55, 73–88.

Jiang, K., Tong, X., Ding, Y.-Q., Wang, Z.-W., Miao, L.-Y., Xiao, Y.-E. et al. (2021) Shifting roles of the East China Sea in the phylogeography of *Machilus thunbergii* (Lauraceae) in East Asia. *Journal of Biogeography*, 2021, 1–16.

Jombart, T. (2008) ADEGENET: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, 24, 1403–1405.

Jones, A.W. & Kennedy, R.S. (2008) Evolution in a tropical archipelago: comparative phylogeography of Philippine fauna and flora reveals complex patterns of colonization and diversification. *Biological Journal of the Linnean Society*, 95, 620–639.

Jones, B.K., Saviola, A.J., Reilly, S.B., Stubbs, A.L., Arida, E., Iskandar, D.T. et al. (2019) Venom composition in a phenotypically variable pit viper (*Trimeresurus insularis*) across the Lesser Sunda Archipelago. *Journal of Proteome Research*, 18, 2206–2220.

Jønsson, K.A., Borregaard, M.K., Carstensen, D.W., Hansen, L.A., Kennedy, J.D., Machac, A. et al. (2017) Biogeography and biotic assembly of Indo-Pacific corvoid passerine birds. *Annual Review of Ecology, Evolution, and Systematics*, 48, 231–253.

Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A. & Jermiin, L.S. (2017) ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods*, 14, 587–589.

Karin, B.R., Stubbs, A.L., Arifin, U., Bloch, L.M., Ramadhan, G., Iskandar, D.T. et al. (2018) The herpetofauna of the Kei Islands (Maluku, Indonesia): Comprehensive report on new and historical collections, biogeographic patterns, conservation concerns, and an annotated checklist of species from Kei Kecil, Kei Besar, Tam, and Kur. *Raffles Bull. Zool.*, 66, 704–738.

Kier, G., Kreft, H., Lee, T.M., Jetz, W., Ibisch, P.L., Nowicki, C. et al. (2009) A global assessment of endemism and species richness across island and mainland regions. *Proceedings of the National Academy of Sciences*, 106, 9322–9327.

Lambert, S.M., Streicher, J.W., Fisher-Reid, M.C., Méndez de la Cruz, F.R., Martínez-Méndez, N., García-Vázquez, U.O. et al. (2019) Inferring introgression using RADseq and DFOIL: power and pitfalls revealed in a case study of spiny lizards (*Sceloporus*). *Molecular Ecology Resources*, 19, 818–837.

Lanfear, R., Frandsen, P.B., Wright, A.M., Senfeld, T. & Calcott, B. (2017) PARTITIONFINDER2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Molecular Biology and Evolution*, 34, 772–773.

Leaché, A.D., Zhu, T., Rannala, B. & Yang, Z. (2019) The spectre of too many species. *Systematic Biology*, 68, 168–181.

Leaché, A.D., Oaks, J.R., Ofori-Boateng, C. & Fujita, M.K. (2020) Comparative phylogeography of West African amphibians and reptiles. *Evolution*, 74, 716–724.

Leaché, A.D., Davis, H.R., Singhal, S., Fujita, M.K., Lahti, M.E. & Zamudio, K.R., (2021) Phylogenomic assessment of biodiversity using a reference-based taxonomy: an example with horned lizards (*Phrynosoma*). *Front. Ecol. Evol.*, 9, 678110

Linkem, C.W., Brown, R.M., Siler, C.D., Evans, B.J., Austin, C.C., Iskandar, D.T. et al. (2013) Stochastic faunal exchanges drive diversification in widespread Wallacean and Pacific island lizards (Squamata: Scincidae: *Lamprolepis smaragdina*). *Journal of Biogeography*, 40, 507–520.

Lohman, D.J., de Bruyn, M., Page, T., von Rintelen, K., Hall, R., Ng, P.K.L. et al. (2011) Biogeography of the Indo-Australian archipelago. *Annual Review of Ecology, Evolution, and Systematics*, 42, 205–226.

Losos, J.B. & Ricklefs, R.E. (2009) Adaptation and diversification on islands. *Nature*, 457, 830–836.

Martin, S.H., Dasmahapatra, K.K., Nadeau, N.J., Salazar, C., Walters, J.R., Simpson, F. et al. (2013) Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. *Genome Res.*, 23, 1817–1828.

Matzke, N.J. (2013) biogeobears: BioGeography with Bayesian (and likelihood) evolutionary analysis in R Scripts. R package, version 0.2.1.

———. (2014) Model selection in historical biogeography reveals that founder-event speciation is a crucial process in island clades. *Systematic Biology*, 63, 951–970.

Mayr, E. (1942) Systematics and the Origin of Species from the Viewpoint of a Zoologist. Harvard University Press, Cambridge, Massachusetts, USA.

Meyer, M. & Kircher, M. (2010) Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor Protocols*, 2010, pdbprot5448.

Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A. et al. (2020) IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Molecular biology and evolution*, 37, 1530–1534.

Monk, K.A., de Fretes, Y. & Reksodiharjo-Lilley, G. (1997) The Ecology of Nusa Tenggara and Maluku. Jakarta, Indonesia: Periplus Editions.

Morinaka, S., Minaka, N., Miyata, T. & Hoshizaki, S. (2017) Phylogeography of the *Delias hyparete* species group (Lepidoptera: Pieridae): complex historical dispersals into and out of Wallacea. *Biological Journal of the Linnean Society*, 121, 576–591.

Moritz, C. (1994) Defining 'evolutionarily significant units' for conservation. *Trends in Ecology & Evolution*, 9, 373–375.

Muraoka, H., Nasution, A., Simanjuntak, J., Dwipa, S., Takahashi, M., Takahashi, H. et al. (2005) Geology and geothermal systems in the Bajawa volcanic rift zone, Flores, Eastern Indonesia. *Proc. World Geotherm. Congr.*, 2005, 1–13.

Mussmann, S.M., Douglas, M.R., Bangs, M.R. & Douglas, M.E. (2020) Comp-D: a program for comprehensive computation of D-statistics and population summaries of reticulated evolution. *Conserv. Genet. Resour.*, 12, 263–267.

Norder, S.J., Proios, K., Whittaker, R.J., Alonso, M.R., Borges, P.A., Borregaard, M.K. et al. (2019) Beyond the Last Glacial Maximum: island endemism is best explained by long-lasting archipelago configurations. *Glob. Ecol. Biogeogr.*, 28, 184–197.

Oaks, J.R., Siler, C.D. & Brown, R.M. (2019) The comparative biogeography of Philippine geckos challenges predictions from a paradigm of climate-driven vicariant diversification across an island archipelago. *Evolution*, 73, 1151–1167.

Ogilvie, H.A., Bouckaert, R.R. & Drummond, A.J. (2017) StarBEAST2 brings faster species tree inference and accurate estimates of substitution rates. *Molecular Biology and Evolution*, 34, 2101–2114.

Oliver, P.M., Blom, M.P.K., Cogger, H.G., Fisher, R.N., Richmond, J.Q. & Woinarski, J.C.Z. (2018) Insular biogeographic origins and high phylogenetic distinctiveness for a recently depleted lizard fauna from Christmas Island, *Australia. Biol. Lett.*, 14, 20170696.

Padial, J.M. & De la Riva, I. (2021) A paradigm shift in our view of species drives current trends in biological classification. *Biol. Rev.*, 96, 731–751.

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

———. (2017) Linking micro- and macroevolutionary perspectives to evaluate the role of Quaternary sea-level oscillations in island diversification. *Evolution*, 71, 2901–2917.

Pease, J.B. & Hahn, M.W. (2015) Detection and polarization of introgression in a five-taxon phylogeny. *Systematic Biology*, 64, 651–662.

Pedersen, M.P., Irestedt, M., Joseph, L., Rahbek, C. & Jønsson, K.A. (2018) Phylogeography of a 'great speciator' (Aves: *Edolisoma tenuirostre*) reveals complex dispersal and diversification dynamics across the Indo-Pacific. *Journal of Biogeography*, 45, 826–837.

Pritchard, J.K., Stephens, M. & Donnelly, P. (2000) Inference of population structure using multilocus genotype data. *Genetics*, 155, 945–959.

Puillandre, N., Lambert, A., Brouillet, S. & Achaz, G. (2012) ABGD, automatic barcode gap discovery for primary species delimitation. *Molecular Ecology*, 21, 1864–1877.

Rambaut, A. & Drummond, A.J. (2009) tracer version 1.7.0. URL http://tree.bio.ed.ac.uk/software/tracer/[accessed 5 July 2020].

Rannala, B. & Yang, Z. (2013) Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. *Genetics*, 164, 1645–1656.

Reilly, S.B., Wogan, G.O.U., Stubbs, A.L., Arida, E., Iskandar, D.T. & McGuire, J.A. (2017) Toxic toad invasion of Wallacea: a biodiversity hotspot characterized by extraordinary endemism. *Glob. Chang. Biol.*, 23, 5029–5031.

Reilly, S.B., Stubbs, A.L., Karin, B.R., Bi, K., Arida, E., Iskandar, D.T. & McGuire, J.A. (2019a) Leap-frog dispersal and mitochondrial introgression: phylogenomics and biogeography of *Limnonectes* fanged frogs in the Lesser Sundas Archipelago of Wallacea. *Journal of Biogeography*, 46, 757–769.

Reilly, S.B., Stubbs, A.L., Karin, B.R., Arida, E., Iskandar, D.T. & McGuire, J.A. (2019b) Recent colonization and expansion through the Lesser Sundas by seven amphibian and reptile species. *Zool. Scr.*, 48, 614–626.

Reilly, S.B., Stubbs, A.L., Arida, E., Arifin, U., Bloch, L., Hamidy, A. et al. (2020) New island records for anurans and squamates from the Lesser Sunda Archipelago. *Herpetol. Rev.*, 51, 785–789.

Reilly, S.B., Stubbs, A.L., Karin, B.R., Arifin, U., Arida, E., Iskandar, D.T. et al. (2021) Genetic divergence of the Sunda ratsnake (*Coelognathus subradiatus*) across the Lesser Sunda Islands (Squamata: Colubridae). *Amphib.-Reptil.*, 42, 269–273.

Reilly, S.B., Stubbs, A.L., Arida, E., Karin, B.R., Arifin, U., Kaiser, H. et al. (2022a) Phylogenomic analysis reveals dispersal-driven speciation and divergence with gene flow in Lesser Sunda flying lizards (Genus *Draco*). *Systematic Biology*, 71, 221–241.

Reilly, S.B., Arifin, U., Stubbs, A.L., Karin, B.R., Kaiser, H., Frederick, J.H. et al. (2022b) Phylogenetic relationships of southern Wallacean ranid frogs (Anura: Ranidae: *Hylarana*). *Zootaxa*, 5150, 591–599.

Ripley, B.D. & Venables, W.N. (2002) Modern Applied Statistics with S. New York, USA: Springer.

Rosauer, D., Laffan, S.W., Crisp, M.D., Donnellan, S.C. & Cook, L.G. (2009) Phylogenetic endemism: a new approach for identifying geographical concentrations of evolutionary history. *Molecular Ecology*, 18, 4061–4072.

Roux, C., Fraïsse, C., Romiguier, J., Anciaux, Y., Galtier, N. & Bierne, N. (2016) Shedding light on the gray zone of speciation along a continuum of genomic divergence. *PLoS Biol.*, 14, e2000234.

Sayyari, E. & Mirarab, S. (2016) Fast coalescent-based computation of local branch support from quartet frequencies. *Molecular Biology and Evolution*, 33, 1654–1668.

Setiadi, M.I., McGuire, J.A., Brown, R.M., Zubairi, M., Iskandar, D.T., Andayani, N. et al. (2011) Adaptive radiation and ecological opportunity in Sulawesi and Philippine fanged frog (*Limnonectes*) communities. *American Naturalist*, 178, 221–240.

Shaffer, H.B. & Thompson, R.C. (2007) Delimiting species in recent radiations. *Systematic Biology*, 56, 896–906.

Shea, G.M. (2012) On the identity of the type species of *Sphenomorphus* (Squamata: Scincidae): *Lygosoma melanopogon* Duméril and Bibron 1839, with a note on a new scalation character of the pes in *Sphenomorphus*. *Zootaxa*, 3490, 1–29.

Simpson, G.G. (1977) Too many lines; the limits of the Oriental and Australian zoogeographic regions. *Proc. Am. Phil. Soc.*, 121, 107–120.

Smith, M.J., Cogger, H., Tiernan, B., Maple, D., Boland, C., Napier, F. et al. (2012) An oceanic island reptile community under threat: the decline of reptiles on Christmas Island, Indian Ocean. *Herpetol. Conserv. Biol.*, 7, 206–218.

Stamatakis, A. (2014) RAXML Version 8: a tool for phylogenetic analysis and postanalysis of large phylogenies. *Bioinformatics*, 30, 1312–1313.

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.

Tallowin, O.J.S., Tamar, K., Meiri, S., Allison, A., Kraus, F., Richards, S.J. & Oliver, P.M. (2018) Early insularity and subsequent mountain uplift were complimentary drivers of diversification in a Melanesian lizard radiation (Gekkonidae: *Cyrtodactylus*). *Molecular Phylogenetics and Evolution*, 125, 29–39.

Tänzler, R., Van Dam, M.H., Toussaint, E.F., Suhardjono, Y.R., Balke, M. & Riedel, A. (2016) Macroevolution of hyperdiverse flightless beetles reflects the complex geological history of the Sunda Arc. *Science Reports*, 6, 18793.

Toussaint, E.F., Müller, C.J., Morinière, J., Tänzler, R. & Balke, M. (2020) A glide over the Indo-Australian geological maze: repeated transgressions of Lydekker's and Wallace's Lines in archdukes, barons and dukes (Nymphalidae: Limenitidinae: Adoliadini). *Biological Journal of the Linnean Society*, 129, 810–821.

Van Dam, M.H. & Matzke, N.J. (2016) Evaluating the influence of connectivity and distance on biogeographical patterns in the southwestern deserts of North America. *Journal of Biogeography*, 43, 1514–1532.

van Welzen, P.C., Parnell, J.A. & Slik, J.F. (2011) Wallace's Line and plant distributions: Two or three phytogeographical areas and where to group Java? *Biological Journal of the Linnean Society*, 103, 531–545.

Vitt, L.J. & Cooper, W.E., Jr. (1986) Skink reproduction and sexual dimorphism: *Eumeces fasciatus* in the southeastern United States, with notes on *Eumeces inexpectatus*. *J. Herpetol.*, 20, 65–76.

White, A.E., Dey, K.K., Stephens, M. & Price, T.D. (2021) Dispersal syndromes drive the formation of biogeographical regions, illustrated by Wallace's *Line. Global Ecol. Biogeogr.*, 30, 685–696.

Whittaker, R.J. & Fernández-Palacios, J.M. (2007) Island Biogeography: Ecology, Evolution, and Conservation. Oxford, United Kingdom: Oxford University Press.

Whittaker, R.J., Fernández-Palacios, J.M., Matthews, T.J., Borregaard, M.K. & Triantis, K.A. (2017) Island biogeography: taking the long view of nature's laboratories. *Science*, 357, eaam8326.

Wiley, E.O. (1978) The Evolutionary Species Concept reconsidered. *Systematic Biology*, 27, 17–26.

Wright, S. (1931) Evolution in Mendelian populations. *Genetics*, 16, 97–159.

Yang, Z. & Rannala, B. (2010) Bayesian species delimitation using multilocus sequence data. *Proceedings of the National Academy of Sciences*, 107, 9264–9269.

Yang, Z. (2015). The BPP program for species tree estimation and species delimitation. *Curr. Zool.*, 61, 854-865.

Zhang, C., Rabiee, M., Sayyari, E. & Mirarab, S. (2018). ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinform.*, 19, 153.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Supplementary Methods: Marker development, exon-capture experiment, and sequence data processing.

Table S1: Sample locality, museum number, and GenBank information.

Table S2: Descriptive statistics for three transcriptomes used in marker development.

Table S3: Dispersal modifiers for the "Stepping-stone" and "PAIC" models

Table S4: Mean uncorrected mitochondrial genome divergence between lineages.

Table S5: Evanno method for determining the optimal number of K from STRUCTURE results.

Table S6: Biogeoghraphical model comparison from BIOGEOBEARS.

Figure S1: *ND4* mitochondrial gene locality map.

Figure S2: Maximum Likelihood phylogeny of the *ND4* gene.

Figure S3: Average target region (exon) sequencing coverage per individual sample.

Figure S4: Average flanking region (intron) sequencing coverage per individual sample.

Figure S5: Sequence data summary plots.

Figure S6: Time-calibrated species tree of the mitochondrial genome from STARBEAST2.

Figure S7: Genomic species trees from ASTRAL-III and STARBEAST2

Figure S8: Ancestral range estimates from BIOGEOBEARS.

Figure S9: PCA of genetic covariance for the Eastern subclade of *S. melanopogon*.

Figure S10: Mantel tests for isolation-by-distance (IBD).

Figure S11: Pruned tree hypotheses used in four-taxon and five-taxon introgression tests.