Local endemism and within-island diversification of shrews illustrate the importance of speciation in building Sundaland mammal diversity

TERRENCE C. DEMOS,*†‡ ANANG S. ACHMADI,§ THOMAS C. GIARLA,*†¶
HERU HANDIKA,**†† MAHARADATUNKAMSI,§ KEVIN C. ROWE**††† and JACOB A. ESSELSTYN*†

*Museum of Natural Science, Louisiana State University, Baton Rouge, LA 70803, USA, †Department of Biological Sciences, Louisiana State University, Baton Rouge, LA 70803, USA, ‡Science and Education, Field Museum of Natural History, Chicago, IL 60605, USA, §Museum Zoologicum Bogoriense, Research Center for Biology-LIPI, Cibinong, Bogor 16911, Indonesia, ¶Department of Biology, Siena College, Loudonville, NY 12211, USA, **Sciences Department, Museum Victoria, Melbourne, Vic. 3001, Australia, ††School of Biosciences, The University of Melbourne, Melbourne, Vic. 3001, Australia

Abstract

Island systems are important models for evolutionary biology because they provide convenient, discrete biogeographic units of study. Continental islands with a history of intermittent dry land connections confound the discrete definitions of islands and have led zoologists to predict (i) little differentiation of terrestrial organisms among continental shelf islands and (ii) extinction, rather than speciation, to be the main cause of differences in community composition among islands. However, few continental island systems have been subjected to well-sampled phylogeographic studies, leaving these biogeographic assumptions of connectivity largely untested. We analysed nine unlinked loci from shrews of the genus *Crocidura* from seven mountains and two lowland localities on the Sundaic continental shelf islands of Sumatra and Java. Coalescent species delimitation strongly supported all currently recognized *Crocidura* species from Sumatra (six species) and Java (five species), as well as one undescribed species endemic to each island. We find that nearly all species of *Crocidura* in the region are endemic to a single island and several of these have their closest relative(s) on the same island. Intra-island genetic divergence among allopatric, conspecific populations is often substantial, perhaps indicating species-level diversity remains underestimated. One recent (Pleistocene) speciation event generated two morphologically distinct, syntopic species on Java, further highlighting the prevalence of within-island diversification. Our results suggest that both between- and within-island speciation processes generated local endemism in Sundaland, supplementing the traditional view that the region's fauna is relictual and primarily governed by extinction.

Keywords: *Crocidura*, island biogeography, Java, phylogeography, speciation, Sumatra

Received 15 September 2015; revision received 9 August 2016; accepted 16 August 2016

Introduction

Islands are appealing natural laboratories of evolution because the surrounding oceans represent obvious barriers for terrestrial species (Wallace 1876, 1881; Schluter 2000; Grant & Grant 2011). However, in continental island systems intermittent dry land connections should reduce isolation, and endemism is anticipated only at the scale of the entire region (Rosenzwieg 1995; Whittaker & Fernández-Palacios 2007). This regional endemism paradigm has led zoologists to predict that (i) terrestrial organisms are widespread within continental island systems, (ii) little evolutionary differentiation

Correspondence: Terrence C. Demos, Fax: (225) 578 3075;
E-mail: terrencedemos@gmail.com

occurs among meta-populations within a system, and (iii) local extinction is the main cause of differences in faunal diversity and composition among continental islands of the same region (MacArthur & Wilson 1967; Heaney 1986; Okie & Brown 2009).

Biologists often view the continental island system of Sundaland (Malay Peninsula, Java, Sumatra, and Borneo) in the context of this regional endemism paradigm (Ruedi 1996; Ruedi & Fumagalli 1996; Gorog et al. 2004; Okie & Brown 2009). This perspective dominates because the assumed recurrent colonization during periods of low sea level should have reduced genetic differentiation among metapopulations (e.g. Heaney 2000; Papadopoulou & Knowles 2015). Recent comparative phylogeographic and phylogenetic studies have explored these issues using mtDNA sequences from limited samples of Sundaic taxa. These studies have generally concluded that Borneo harbours more distinctive lineages than other Sundaic islands (de Bruyn et al. 2014; Leonard et al. 2015; Sheldon et al. 2015). However, many of these studies largely exclude Javan lineages because of a lack of samples. Furthermore, Bornean material is often dominated by specimens from Malaysia, with little or no material from Kalimantan, which represents 73% of the island's land area.

Several authors have invoked extinction to explain the differences in diversity and composition among vertebrate communities of Sundaic islands (Heaney 1986; Okie & Brown 2009; Wilting et al. 2012; den Tex & Leonard 2013). These interpretations reinforce the regional endemism paradigm (e.g. Brown 1986) and implicate local extinction as the primary generator of β-diversity within continental island systems. Alternatively however, at least some of the pattern could be explained by between-island diversification, especially if it has occurred between the larger islands. In essence, either extinction or speciation can generate a species distribution that covers only a portion of a continental island system, but absences of Sundaic species on particular islands have traditionally been interpreted as extinctions. While this interpretation is certainly true for many examples (Piper et al. 2008; Cranbrook 2010), some studies have found relatively deep mtDNA divergences among populations both between and within large Sundaic islands (Gorog et al. 2004; Esselstyn et al. 2010; Oliveros & Moyle 2010; Roberts et al. 2011), raising the idea that speciation may generate β-diversity on the Sunda shelf. If isolating mechanisms at the intra-island scale are sufficient to generate speciation, then mechanisms at the larger, between-island scale also are almost certainly sufficient to produce the same effect. As such, densely sampled, fine-scale phylogeographic studies may be more informative than sparsely sampled, broad-scale studies at determining the relative importance of speciation in generating β-diversity among continental islands.

Terrestrial vertebrate species are not fully documented for many Sundaland taxa, further obfuscating the historical formation of the region's biota. Even 'well-studied' groups such as mammals are incompletely known, as demonstrated by recent discoveries of new species on the Sunda Shelf (Achmadi et al. 2012; Esselstyn et al. 2014) and neighbouring areas (Heaney et al. 2011; Esselstyn et al. 2012, 2015; Rowe et al. 2016). Tropical regions often contain a glut of data-deficient (DD) species (IUCN 2015). Those areas that are rich in DD taxa also tend to harbour a disproportionate number of species that were recently described (Brito 2010) or have narrow geographic distributions (Sheth et al. 2012). Among mammals, Sundaland is a 'hotspot' of DD species (Bland et al. 2015), and hence, one might expect many species and range extensions to await discovery. These deficiencies in knowledge of Sundaland species and their distributions may have biased biogeographic inferences (e.g. Heaney 2007; Esselstyn et al. 2010; Stelbrink et al. 2012; de Bruyn et al. 2014; Leonard et al. 2015; Merckx et al. 2015).

Shrews in the genus Crocidura are prevalent members of small mammal communities in Sundaland. However, because of the cryptic nature of morphological diversity in Crocidura, and the lack of adequate comparative series, authors have often disagreed on the number and composition of species in Sundaland. For example, a series of morphological revisions and faunal summaries (Jenkins 1982; Corbet & Hill 1992; Ruedi 1995) recognized 2–6 species of Sumatran Crocidura, with 0–5 of them regarded as endemic. The systematics of Javan shrews is somewhat better resolved, having been the subject of recent molecular and morphological investigations (Esselstyn et al. 2013, 2014), but some taxonomic issues remain (see Materials and methods).

In this study, we used DNA sequences from nine unlinked loci to (i) estimate species boundaries and population structure of shrews within and between the islands of Sumatra and Java, (ii) place these species in a broad phylogenetic context and (iii) assess the geographic scale of endemism among the shrews of Sumatra and Java.

Materials and methods

Species sampling and study area

We sampled all 11 currently recognized species of Sumatran and Javan Crocidura, as well as two putative undescribed species (224 specimens total). Geographically, our sampling is derived from inventories of shrew species from at least one site on each of five

Javan mountains (from west to east: Mts. Salak, Gede, Ciremai, Slamet and Ijen); two Sumatran mountains (Mts. Singgalang and Tujuh); and two lowland sites on (Mt. Leuser National Park), and adjacent (Bangka Island) to, Sumatra (Fig. 1). All sample sites were in forested or forest edge habitats. Our phylogenetic analyses included an additional 17 South-East Asian *Crocidura* species (1–2 samples each) and an African out-group, *Crocidura monax*. Hence, our total sampling includes 263 specimens representing 28 recognized and two putative species. This includes 19 of 20 species known from Sundaland and 28 of 49 species known from South-East Asia east of the Thailand–Myanmar border and south of the Ryukyu Islands, including the Philippines and Sulawesi (Jenkins *et al.* 2009, 2010, 2013; Esselstyn & Goodman 2010; Esselstyn *et al.* 2010, 2014 [and references therein]; Appendix S1, Supporting information). Preliminary identification to species was made using morphological characters described in Jenkins (1982) and Ruedi (1995) and later refined by examining

mtDNA gene tree topology, with subsequent re-examination of morphology.

Taxonomy of Javan and Sumatran Crocidura

Sumatran *Crocidura* can be provisionally grouped on the basis of body size. *Crocidura neglecta* is much smaller (<5 g) than other Sumatran species and was until recently (Esselstyn *et al.* 2013) known only from the holotype. Ruedi (1995) included *C. neglecta* in the widespread Sundaland *C. monticola* complex, but Esselstyn *et al.* (2013) found it to be a distant relative of *C. monticola* from Java (the type locality). Ruedi (1995) indicated that the other small Sumatran species, *C. beccarii* (5–8 g), may be a close relative of *C. vosmaeri* (5.8–8.8 g), a possible endemic to Bangka Island (Fig. 1). Among medium-sized species, Ruedi (1995) described *C. hutanis* (10–12 g) and recognized the long-tailed species *C. paradoxura* as a Sumatran endemic. Finally, the relatively large *C. lepidura* (13–21 g) was assigned to
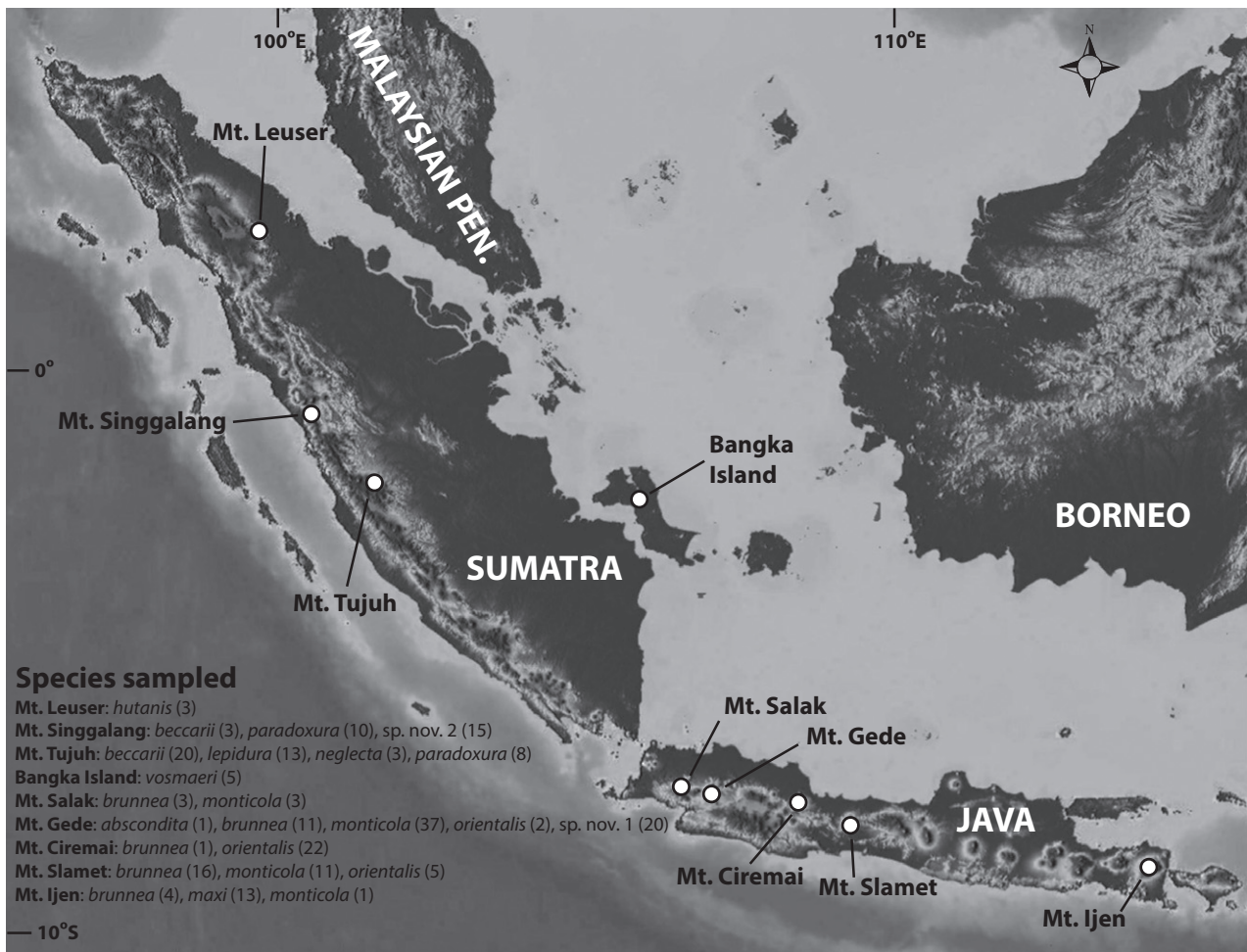

Fig. 1 Map of sampling localities on Sumatra and Java. For each location sampled, we list the species followed in parentheses by the number of individuals of *Crocidura* that we sequenced. Catalog numbers are given in Appendix S1 (Supporting information).

the geographically widespread *C. fuliginosa* complex by Jenkins (1982) and Corbet & Hill (1992). Ruedi (1995), however, considered *C. lepidura* a Sumatran endemic. Among Javan shrews, Esselstyn *et al.* (2013, 2014) recognized the large *C. orientalis* and *C. brunnea*, the small *C. monticola* and *C. maxi* and the newly discovered *C. abscondita* (mistakenly named *C. absconditus*, which uses the incorrect gender). In the light of new specimens of true *C. maxi* from East Java, we now treat the putative *C. maxi* from West Java, reported by Esselstyn *et al.* (2013), as an undescribed species. Genetic data (see Results) support this conclusion and render the taxonomy outlined in this study more consistent with previous studies (e.g. Kitchener *et al.* 1994).

Molecular methods

Specimens were sequenced for a portion of the mitochondrial cytochrome-b (*cyt-b*) and eight unlinked nuclear genes, including seven exons (*ApoB*, *BDNF*, *BRCA1*, *GHR10*, *PTGER4*, *RAG1* and *vWF*) and one intron (*MCGF*). Methods of DNA extraction, PCR and sequencing follow those of Esselstyn *et al.* (2009, 2013). Chromatographs were checked manually, assembled and edited using GENEIOUS PRO 7.1.7 (Biomatters Ltd.). Newly generated sequences were deposited in GenBank (KX469457–KX470389; Appendix S1, Supporting information). Sequences from each locus were aligned independently using the MUSCLE algorithm (Edgar 2004) with default settings in GENEIOUS. Sequence data from *cyt-b* and the seven exons were translated into amino acids and inspected for deletions, insertions and premature stop codons to prevent inclusion of paralogous sequences. Alignments for all data sets were inspected visually and determined to be unambiguous. Nuclear alleles were resolved statistically using PHASE 2.1 (Stephens *et al.* 2001) under default parameters, except that we adjusted the haplotype acceptance threshold to 0.70, which has been shown to reduce the number of unresolved genotypes with little to no increase in false positives (Garrick *et al.* 2010). Input files for PHASE were assembled using the SEQPHASE web server (Flot 2010). PHASE was run for 1000 iterations with a burn-in of 500 and a thinning interval of 1. We tested for recombination using the Detect Recombination plugin in GENEIOUS.

Substitution models and gene tree estimation

We used the Bayesian information criterion (BIC), as implemented in PartitionFinder (Lanfear *et al.* 2012), to identify the optimal partitioning scheme and best model of nucleotide substitution for each partition in the *cyt-b* alignment. The most appropriate model of evolution for each unpartitioned nuclear gene was determined using the BIC on the maximum-likelihood topology estimated for each model independently in JMODELTEST v.2.1.7 (Darriba *et al.* 2012). We used the greedy search algorithm and linked branch lengths for likelihood score calculations in JMODELTEST. Gene trees were inferred using maximum-likelihood and Bayesian methods for *cyt-b* and for individual, phased nuclear genes. Maximum-likelihood estimates of gene trees were made in GARLI v.2.1 (Zwickl 2006) using default settings and 1000 bootstrap replicates. GARLI runs were replicated five times for the *cyt-b* locus and 100 times for the nuclear loci to ensure consistency. The tree that received the highest likelihood was reported for each analysis, and bootstrap scores were summarized on these ML trees using SumTrees in DENDROPY (Sukumaran & Holder 2010). Bayesian gene tree analyses used MRBAYES v.3.2.5 (Ronquist *et al.* 2012) and two replicates were run to ensure proper mixing had occurred. Nucleotide substitution models were unlinked across partitions and were allowed to evolve at individual rates in the *cyt-b* locus. Eight Markov chains with default heating values were conducted for 5×10^6 generations and sampled every 1000th generation. Stationarity was assessed using TRACER v.1.6 (Rambaut *et al.* 2014). The first 1000 samples were discarded as burn-in and the remaining 4000 samples formed the posterior probability (PP) distributions. Majority rule consensus trees were generated from each analysis.

Population structure

We clustered nuclear alleles using STRUCTURE v.2.3.4 (Pritchard *et al.* 2000; Hubisz *et al.* 2009) to infer population-level diversity. Our goal was to determine whether assignment of individuals to populations was consistent with (i) clade membership inferred from the mtDNA gene tree, (ii) the geographic origin of samples and (iii) morphology-based species identifications. Mitochondrial DNA was excluded to avoid circularity.

We carried out a hierarchical series of STRUCTURE analyses for Sumatran and Javan species groups, as well as individual species within each group. First, analyses were conducted independently on six subsets of taxa, where uncertainty in species limits was suggested by earlier morphological (Jenkins 1982; Ruedi 1995) or molecular studies (Esselstyn *et al.* 2009, 2013; Omar *et al.* 2013), or because recently collected specimens could not be assigned to any currently recognized taxon. These sets were composed of (i) *Crocidura monticola* (Java), *C.* sp. nov. 1 [Java (*C. maxi* in Esselstyn *et al.* 2013)] and *C.* sp. nov. 2 (Sumatra); (ii) *C. beccarii* (Sumatra), *C. vosmaeri* (Bangka), *C. lepidura* (Sumatra) and *C. hutanis* (Sumatra); (iii) *C. orientalis* (Java); (iv) *C. brunnea* (Java); (v) *C. maxi* (Java and Lesser Sundas);

and (vi) *C. paradoxura* (Sumatra). Next, individual STRUCTURE analyses were employed for each of the aforementioned species, excluding those sampled from only a single location, in which case they were analysed with their putative sister species, as inferred on the rooted mtDNA gene tree. We used the admixture model with correlated allele frequencies to allow for mixed ancestry of individuals. The number of clusters (K) was inferred using 10 replicates for each K value with a burn-in of 2×10^4–10^5 and 2–5×10^5 iterations. The maximum value of K (K_{max}) for each analysis was calculated as one more than the sum of the number of localities sampled per species. Two independent runs of 10 replicates were conducted for each pooled set of individuals at each K between 1 and K_{max}. The optimal value of K was determined using the ΔK method of Evanno *et al.* (2005), implemented on the CLUMPAK web server (Kopelman *et al.* 2015). However, the ΔK method may underestimate the optimal number of clusters in the presence of hierarchical structure (Waples & Gaggiotti 2006), and has been shown to fail to recover the true value of K when subpopulation sample sizes are small (<10) and $K > 2$, for example (Gao *et al.* 2011). Therefore, we also heuristically examined the differences in log-likelihood values among simulations to exclude ΔK-supported values that were biologically unrealistic (i.e. in conflict with a combination of phylogenetic inference, morphology and geographic distributions *sensu* Meirmans 2015). Cluster membership probabilities were provided by CLUMPP (Jakobsson & Rosenberg 2007), and results were visualized using DISTRUCT (Rosenberg 2004), also on the CLUMPAK server.

Lineage delimitation

We conducted joint independent lineage delimitation and species tree estimation using the program BPP v.3.1 (Yang & Rannala 2010, 2014). Independent BPP analyses were carried out on the same six subsets of taxa described above for STRUCTURE analyses. Within each of these subsets, we designated each species from each locality as a putative independent lineage, effectively putting a maximum on the number of lineages that could be delimited. While it is a widespread practice to use BPP to explicitly delimit species under a unified lineage species concept (de Queiroz 2007), we refrain from using BPP to formally diagnose species in this study because of our inability to assess the possible influence of isolation by distance on delimitation analyses. Assigning localities as putative populations is a conservative approach that consistently recovers the same number of genetically isolated populations in BPP as using a priori population assignment based on independent analyses (Leaché & Fujita 2010; Camargo *et al.*

2012; Demos *et al.* 2014b). We tested the validity of our assignment of individuals to morphospecies using both a guide tree generated from the multilocus species tree inferred using *BEAST and the guide-tree-free implementation of BPP. Initial analyses showed that algorithms 0 and 1 (Yang & Rannala 2010) produced similar results; therefore, algorithm 0 was implemented for subsequent analysis. We used initial tuning values and Γ shape parameters chosen by Giarla *et al.* (2014) and trial runs showed good mixing. For each of the five data sets, we used all eight phased nuclear loci. Two Γ-distributed prior probability schemes were used to compare the effects of large and small populations sizes ($\theta = \Gamma$ [1, 10] and Γ [2, 2000], respectively) on delimitation results. The divergence time prior, τ, used a diffuse Γ-distributed probability distribution Γ (2, 2000), with a mean of 0.001, which assumes that species split one million years ago if substitution rates are 2.2×10^{-9} (Kumar & Subramanian 2002) and generation time is equal to 1 year. All BPP analyses were run for 10^6 generations, with a burn-in of 10^5 generations and samples were drawn every fifth generation. We carried out a replicated analysis of each data set to ensure convergence and proper mixing of the rjMCMC algorithm. Thus, 10 BPP runs were conducted for each of the six data sets. To ensure that BPP was not arbitrarily delimiting incorrect groups, we randomized individual assignments to populations once, and ran BPP analyses following the procedure of Burbrink *et al.* (2011).

Species tree estimation

We estimated a species tree in *BEAST 2.1.1 (Drummond *et al.* 2012) using the eight nuclear (nDNA) alignments and all species of Javan and Sumatran *Crocidura*, plus a broader sample of SE Asian *Crocidura.* For those Javan and Sumatran species for which samples from more than one disjunct population exist, we assigned samples from separate localities as terminal taxa, resulting in 42 tips in these analyses. The nDNA loci were reduced to three individuals per species or population when $n > 3$ to keep analyses tractable and facilitate convergence. During initial runs, nucleotide substitution models selected using JMODELTEST were applied to individual loci; however, difficulty in achieving proper MCMC mixing necessitated the use of simpler models. We therefore adopted HKY models that did not include Γ-distributed rate parameters or proportion of invariant site parameters for six loci. The substitution, clock and tree models were unlinked across all loci. The uncorrelated lognormal relaxed clock was applied to each locus with a Yule tree prior and the constant root population size model. Four replicate analyses were conducted with random starting seeds and chain lengths of

2×10^9 generations, with parameters sampled every 2×10^5 steps. Long chains were necessary for achieving high effective sample sizes (ESS) for parameters. Convergence was assessed in TRACER v.1.6 (Rambaut *et al.* 2014). The first 25% of trees were removed as burn-in, and the maximum clade credibility tree was assembled using LOGCOMBINER v.2.1.1 and TREEANNOTATOR v.2.0.3 (Drummond *et al.* 2012).

Estimating interspecific gene flow

Assessing gene flow, or a lack thereof, can indicate the strength of putative ecological barriers and whether lineages should be treated as independent species. Therefore, we used a model-testing framework implemented in IMA2 (Hey & Nielsen 2007; Hey 2010) to compare models of divergence history with and without gene flow. We analysed two pairs of species from Java (i) *Crocidura brunnea* and *C. orientalis* that appear to be elevationally partitioned on Mt. Slamet; and (ii) *C. monticola* and an undescribed species (*C. maxi* in Esselstyn *et al.* 2013) that co-occur at mid elevations on Mt. Gede (Esselstyn *et al.* 2013). We estimated the joint PP of the migration parameters m_1 and m_2 for populations of the above species pairs using our complete phased nDNA data set and applied the HKY model for all genes. We performed extensive preliminary runs to identify appropriate bounds on demographic parameter priors and to optimize the MCMC settings for sufficient mixing. Mixing was assessed by inspection of ESS, parameter trend plots, and update rates. The recording phase for both species pairs included 30 independent Markov chains for 10^6 steps sampled every 10 steps with a burnin of 2×10^5. For *C. monticola* versus *C.* sp. nov. 1, the upper prior limits were $q = 3$, $t = 8$, $m = 10$. For *C. brunnea* vs. *C. orientalis*, the upper prior limits were $q = 3$, $t = 8$, $m = 3$. Both species pair analyses used a geometric heating scheme (-h$_a$ = 0.96 -h$_b$ = 0.90). Two independent M-mode runs with different starting seeds were performed for each species-pair analysis. We used the L-mode analyses to compare four nested migration models against the full migration model: (i) individual coalescent migration rates for species 0 and species 1; (ii) a single coalescent migration rate for both species; (iii) no migration from species 0 to species 1; (iv) no migration from species 1 to 0; and (v) an isolation model with no migration (cf. Kerhoulas *et al.* 2015). From this output, nested models were ranked by relative Akaike information criterion (AIC) differences among models using $-\log(P)$ values from the L-mode analyses as described in Carstens *et al.* (2009). Following Carstens *et al.* (2009), we also calculated Akaike weights (ω_i, normalized relative model likelihoods) and the evidence ratio ($E_{\min}/i$, a comparison of each model to the best

model as an objective measure of model support) as additional measures of model support. An evidence ratio of <10 can be considered as moderate support for a model relative to the best model (Burnham & Anderson 2002).

Results

Loci, taxon sampling and sequence alignment

Our *cyt-b* (1110 bp) alignment contained 245 individuals (full or partial coverage), 92 of which were newly generated. The alignment includes 450 variable sites, 410 of which are parsimony informative. To aid in visualization of phylogenies inferred from this matrix, we reduced the matrix of 245 individuals to a set of unique sequences, resulting in a final alignment of 113 haplotypes. Complete nDNA alignments (4421 bp total) for use in individual nuclear gene tree analyses contained 476–500 alleles for each gene (Figs S1 and S2, Appendix S1, Supporting information; Dryad doi: 10.5061/dryad.362pt). Overall, ~4% of data were missing for the 8 nDNA loci (Appendix S1, Supporting information). The reduced nDNA data set for species tree inference contained 98 individuals and 176–190 phased sequences per gene, each with 23–126 variable sites and 19–117 parsimony informative sites. We found no evidence of intralocus recombination from the four-gamete tests.

Phylogenetic relationships

The *cyt-b* gene trees generated by MrBayes and GARLI contained strong support for many nodes, but those surrounded by short branches or in deeper parts of the tree tended to receive limited support (Fig. 2). Javan and Sumatran species or clades are dispersed across the topology with four clades from each island. Interisland phylogenetic relationships include the following: *Crocidura maxi* from Java is strongly supported as sister to *C. maxi* populations from the Lesser Sunda Islands and Bali and this multi-island *C. maxi* clade is strongly supported as sister to *C. elongata* from Sulawesi; the Javan endemics *C. orientalis* and *C. brunnea* are sisters and together are strongly supported as sister to four species endemic to Sumatra; another Javan endemic, *C. abscondita*, is poorly supported as sister to *C. negligens* from peninsular Malaysia; and finally, a well-supported clade that includes two Javan endemic species (*C. monticola* and *C.* sp. nov. 1) is a member of a polytomy with species from throughout South-East Asia and Sumatra. A well-supported clade of *C. maxi* whose island distributions straddle Wallace's Line was inferred (populations from Aru + Alor + Java Islands). *Crocidura monticola* is inferred to be paraphyletic, with *C.* sp. nov. 1 forming a
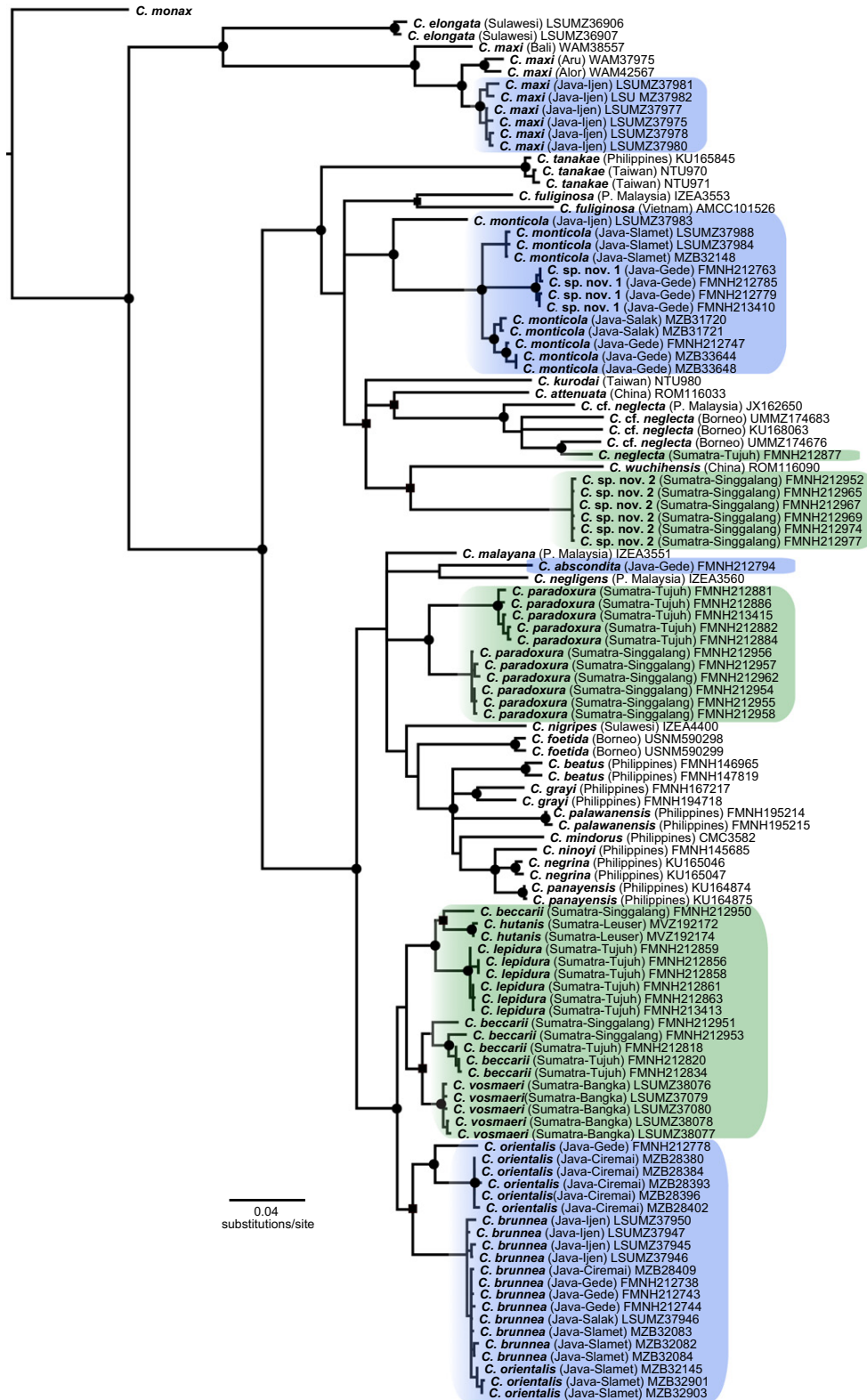
Fig. 2 Cytochrome-b gene tree inferred using maximum-likelihood and Bayesian inference in the programs GARLI and MrBayes, respectively. Clades distributed on Sumatra are highlighted with green and clades distributed on Java are highlighted with blue. Black circles on nodes indicate ML bootstrap ≥0.70 and Bayesian posterior probability (PP) ≥0.95. Black squares indicate ML bootstrap ≥0.50 and <0.70, and PP ≥0.75 and <0.95. Nodes with ML bootstrap <0.50 and PP <0.75 are not marked.

polytomy with two clades of *C. monticola*, which together are sister to another population of *C. monticola*. From Sumatra, we also recovered four clades that are consistent with multiple origins for Sumatran species. At the interisland/island–mainland level, the following relationships are recovered: *Crocidura neglecta* from Sumatra is strongly supported as sister to *C.* cf. *neglecta* from Borneo (formerly *C.* cf. *monticola*; Ruedi 1995); *C.* sp. nov. 2 (Sumatra) is weakly supported as sister to *C. wuchihensis* from China; *C. paradoxura* is a member of a polytomy that includes multiple species from throughout South-East Asia; and a Sumatran clade that includes four species is strongly supported as sister to a Javan clade consisting of *C. brunnea* and *C. orientalis*. A within-Sumatra multi-species clade is recovered that includes *C. beccarii*, *C. hutanis*, *C. lepidura* and *C. vosmaeri*. In total, at least five intra-island (in situ) speciation events are inferred, two for Java and three for Sumatra.

Evidence of possible introgressive hybridization is evident based on incongruence of mtDNA gene trees and morphology for samples of *Crocidura orientalis* from Mt. Slamet on Java that are more closely related to *C. brunnea* than to two additional *C. orientalis* lineages from Mts. Ciremai and Gede. Both of these species are morphologically diagnosable with external characters (Ruedi 1995; Esselstyn *et al.* 2014). While sympatric populations of *C. orientalis* and *C. brunnea* are also distributed on Mts. Ciremai and Gede, those samples sort to their respective species-level clades. In addition, one sample of *C. beccarii* from Mt. Singgalang, Sumatra, was recovered as a member of the *C. hutanis* lineage from lowland forest in Mt. Leuser NP, Sumatra. There were no other mtDNA haplotypes shared among localities.

ESS for all but two parameters exceeded 200 in the species tree analysis. The exceptions were the tree likelihoods for *ApoB* and *Rag1*, which were each >100. Phylogenetic relationships inferred in our *BEAST analysis generally agreed with the mtDNA gene trees in their support for topological relationships between the 13 Javan and Sumatran species included in our analyses (Fig. 3). Five separate Javan and/or Sumatran clades are inferred in the species tree while seven are inferred in the mitochondrial gene trees. Species tree phylogenetic estimates support *Crocidura brunnea*, *C. orientalis* and *C. maxi* from Java and *C.* sp. nov. 2, *C. neglecta*, *C. paradoxura*, *C. hutanis* and *C. lepidura* from Sumatra as monophyletic. *Crocidura* sp. nov. 1 from Mt. Gede, Java, is nested within the four *C. monticola* populations distributed across Java and is sister to *C. monticola* from Mt. Ijen, the most distant Javan sample site (Fig. 1). *Crocidura vosmaeri* from Bangka Island, just off Sumatra, is nested within *C. beccarii* where it is sister to the Mt. Singgalang population.

Population structure

We carried out STRUCTURE analyses to test for differentiation between (i) *Crocidura beccarii* and *C. vosmaeri*, (ii) *C. lepidura* and *C. hutanis* and (iii) the two putative new species and their respective sister species/clades. All six of these putative species are supported as distinctive using the delta *K* method (Fig. 4 and Table 1; Evanno *et al.* 2005). We also tested allopatric populations of individual species for isolation. The two sampled populations of *C. paradoxura* from Sumatra also were distinguished by STRUCTURE with minimal evidence of admixture. The Javan species with samples available from more than one population (i.e. *C. monticola*, *C. brunnea* and *C. orientalis*) exhibit varying degrees of population structure. In none of these three species was each population assigned to a separate cluster (Fig. 4 and Table 1). The potentially widespread species of *C. maxi* from eastern Java and the Lesser Sunda Islands, and *C. neglecta* from Sumatra and Borneo, had best-supported *K* values of two using the Evanno method (Fig. S3, Supporting information). However, examination of ancestry proportions using DISTRUCT revealed no population structure and the likelihood of models for *K* = 1 in STRUCTURE were the highest among the models tested (*K* = 1–4).

Coalescent delimitation

Coalescent analyses in BPP that treated each isolated sample location as a potential species supported delimitation of 20–22 lineages among the 13 species we recognized from morphology and the mtDNA gene tree topology. These delimitation results were minimally affected by varying the prior distributions on mutation rate-scaled effective population sizes (θ) and divergence times (τ_0). The combination of large ancestral population sizes and shallow divergences resulted in marginally lower support values for the populations and species to which they were assigned (Table 1). We considered any PP ≥0.99 for any guide tree or prior scheme as strong support for a putative speciation event (Table 1). Randomization of individuals into clades resulted in the collapse of all nodes that previously had PP ≥ 0.99, indicating BPP is not simply delimiting all lineages. The joint estimation of guide trees and delimitation by BPP vs. implementation of the *BEAST generated guide tree for delimitation resulted in modest variation in posterior probabilities for speciation and one fewer delimited species in the latter set of analyses (i.e. *Crocidura brunnea* from Salak). All of the previously named Sumatran lineages tested using BPP were distinct with posterior probabilities of 1.0. Contrary to the results
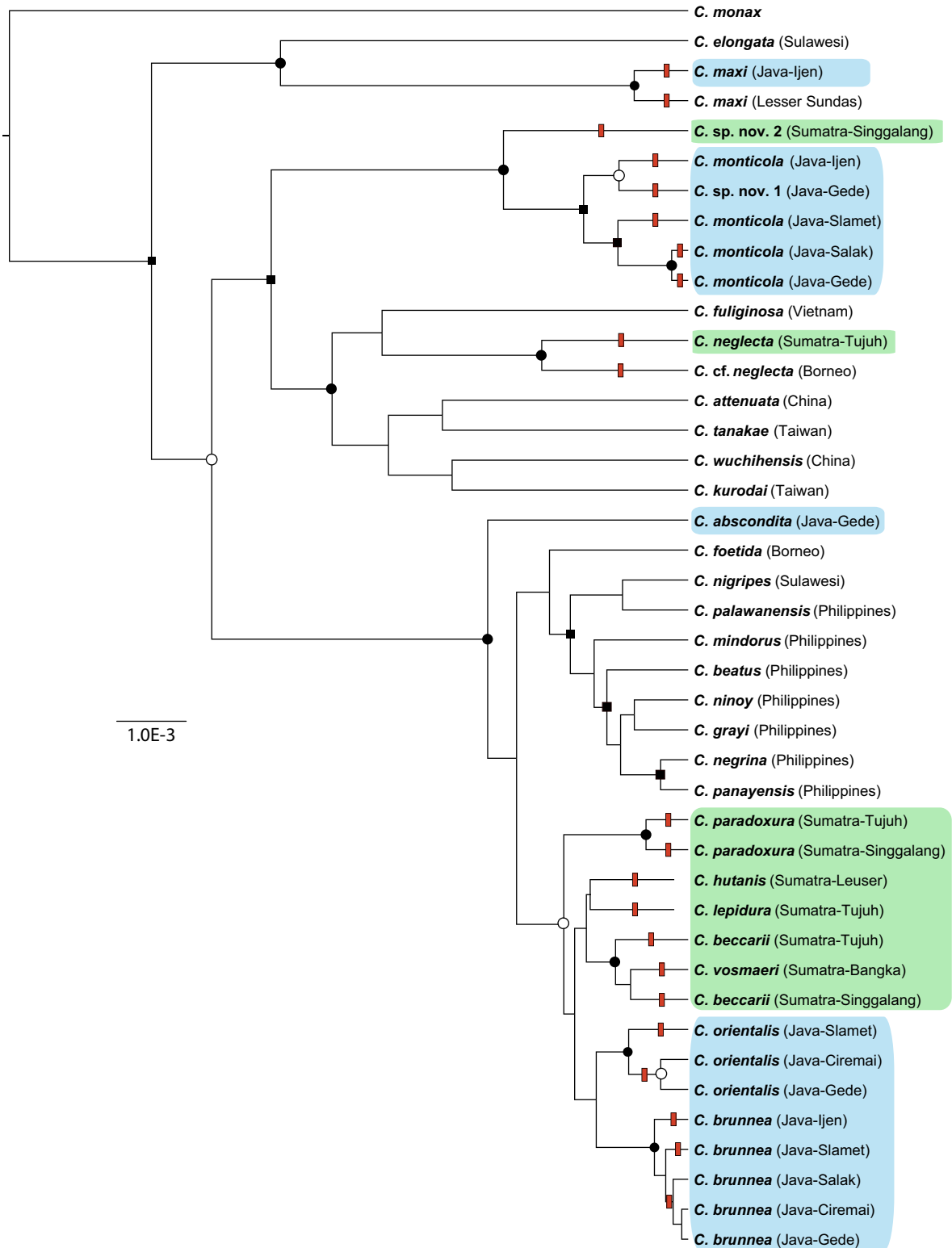
Fig. 3 Multilocus species tree inferred using **beast*. Posterior probabilities are indicated by filled circles if ≥0.95, filled squares if ≥0.85 and <0.95, and open circles if ≥0.70 and <0.85. Nodes with posterior probability (PP) <0.70 are not marked. Terminals are labelled with species names followed by region of origin in parentheses. Javan species and populations are highlighted in blue and Sumatran species and populations are highlighted in green. Red rectangular bars bisecting branches indicate results from bpp species delimitation analyses with posterior probabilities ≥0.99 for a given lineage.
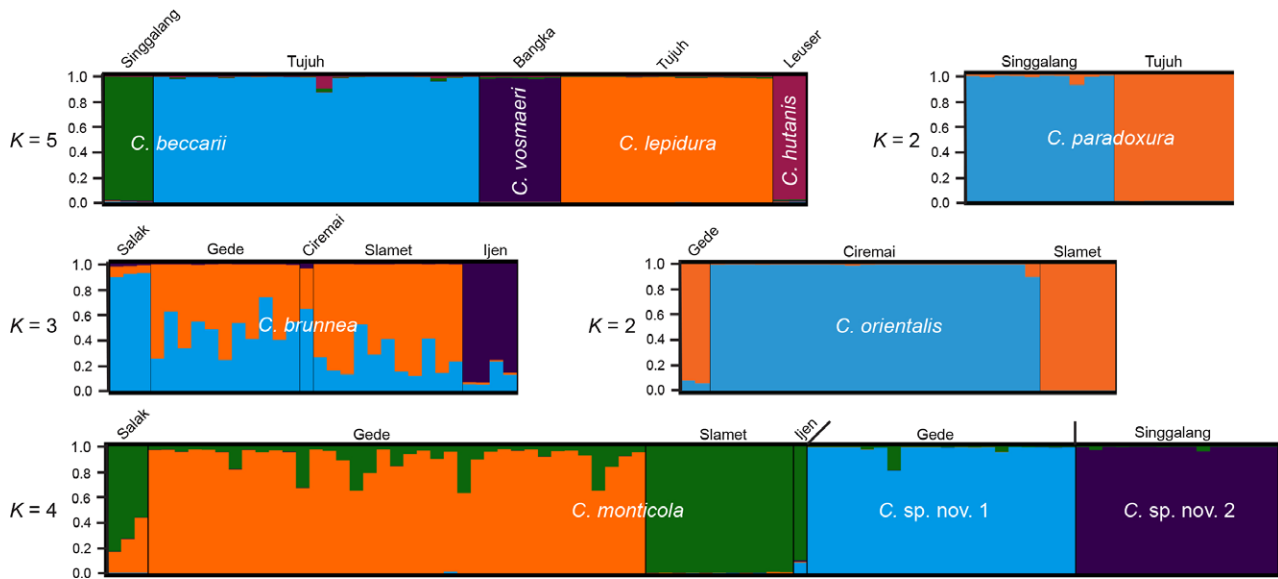
Fig. 4 DISTRUCT visualization of STRUCTURE analyses assigning individuals to major populations for Sumatran and Javan *Crocidura*.

from STRUCTURE, *C. neglecta* from Sumatra was delimited from its sister lineage on Borneo. The total number of putative species on Sumatra suggested by BPP was nine, an increase of three over our initial morphological and mtDNA conclusions.

Among the Javan samples analysed in BPP, we recovered strong support (PP = 1.0) to delimit 11 lineages on Java. The only previous molecular phylogenetic assessment of Javan *Crocidura* diversity based on fewer sample localities found strong support for six species using the same method (Esselstyn *et al.* 2013). Despite incomplete lineage sorting or potential introgression of mitochondrial and nuclear loci between *Crocidura brunnea* and *C. orientalis* on Mt. Slamet, we delimited lineages with strong support within each of these species (Fig. 3, Table 1). We also delimited populations of *C. maxi* from eastern Java (Mt. Ijen) from non-Sunda Shelf populations from the Lesser Sunda Islands with strong support. Finally, we delimited, with a probability of 1.0 in all BPP analyses, the putative undescribed species (*C.* sp. nov. 1) from Mt. Gede (incorrectly referred to *C. maxi* by Esselstyn *et al.* 2013) from a poorly supported sister lineage of *C. monticola* from Mt. Ijen. The clade that includes all populations of *C. monticola* + *C.* sp. nov. 1 from Java is well supported as sister to *C.* sp. nov. 2 from Sumatra and all populations within this clade are delimited with high support. There is minimal sequence divergence between *C. monticola* and *C.* sp. nov. 1 (3.8% mtDNA uncorrected *p*-distance). However, in this case, several external phenotypic characters make the two species (i.e. *C. monticola* as currently described, and the putative *C.* sp. nov. 1) diagnosable and support their status as distinct species (Esselstyn *et al.* 2014). We did not use BPP

to test one recently described species from Java, *C. abscondita*, as it is distantly related to other species in both gene trees and the species tree. Based on available samples, it has no close relatives and its known distribution is restricted to Mt. Gede.

Gene flow

Results from IMA2 were nearly identical between independent runs, and therefore, we present results from only one run. For the species pair on Mt. Slamet, the best-supported model was unidirectional migration from *C. brunnea* into *C. orientalis* based on the AIC scores of ranked models from the L-mode analysis in IMA2 (Table S1, Supporting information). The model of no migration (pure isolation model) was the least supported. Our analyses indicated significant unidirectional gene flow from *C. brunnea* into *C. orientalis* on Mt. Slamet [Table 2; log-likelihood ratio test (LLR) = 123.250, *P* < 0.001; posterior distribution peak at 0.21 migrants per generation]. There were also indications of hybridization in one *C. orientalis* specimen based on shared alleles at three nDNA loci (Figs S1 and S2, Supporting information). Gene flow was near zero and not significant from *C. orientalis* into *C. brunnea* (Table 2; LRR = 0, n.s.). For the species pair on Mt. Gede, the best-supported model based on ranked AIC scores was unidirectional migration from *C.* sp. nov. 1 into *C. monticola* based on the ranked models from the L-mode analysis (Table S2, Supporting information). The model of no migration was the least supported model. Analyses of these putative sibling species indicated significant unidirectional gene flow from *C.* sp. nov. 1 into *C. monticola* (Table 2; LLR = 33.334,

Table 1 Summary of results from BPP3 and STRUCTURE analyses. Prior schemes tested in BPP indicated by BPPsm and BPPlr. In both schemes, divergence depths assuming a mutation rate of 10^{-9} substitutions per site per year (Kumar & Subramanian 2002) indicate a divergence depth prior with a mean of ~1.0 Ma. BPPlr indicates a θ prior of ~0.1 and BPPsm indicates a θ prior of ~0.001. The guide tree was based on the *BEAST species tree for the first pair of analyses (species tree guide tree). Species tree inference and species delimitation were jointly inferred for the second pair of analyses without a priori provision of a guide tree (BPP guide tree free). STRUCTURE admixture indicates that DISTRUCT plots for populations in which more than one individual had a partial assignment to more than one group ($q < 0.90$) were considered admixed. STRUCTURE assignment indicates the population or combined populations that individuals were assigned to (numbers assigned to populations in left of Table 1). Agree indicates that BPP3 and STRUCTURE analyses supported the same populations as genetically isolated. Monophyly indicates the *cyt-b* tree supports reciprocal monophyly for a clade

			Species tree guide tree		BPP guide tree free		STRUCTURE admixture	STRUCTURE assignment	Agree	Monophyly
			BPPlr	BPPsm	BPPlr	BPPsm				
Java	1	*C. brunnea* (Ciremai)	0.87	0.99	0.77	0.55	Admixed	1 + 2 + 5	—	No
	2	*C. brunnea* (Gede)	0.82	0.99	0.77	0.55	Admixed	1 + 2 + 5	—	No
	3	*C. brunnea* (Ijen)	1.0	1.0	1.0	1.0	Admixed	3	—	No
	4	*C. brunnea* (Salak)	0.82	0.99	0.99	0.99	—	4	—	No
	5	*C. brunnea* (Slamet)	1.0	1.0	1.0	1.0	Admixed	1 + 2 + 5	—	No
	6	*C. orientalis* (Ciremai)	1.0	1.0	0.98	1.0	—	6	Yes	Yes
	7	*C. orientalis* (Gede)	1.0	1.0	0.98	1.0	—	7 + 8	—	Yes
	8	*C. orientalis* (Slamet)	1.0	1.0	1.0	1.0	—	7 + 8	—	Yes
	9	*C. monticola* (Gede)	1.0	1.0	1.0	1.0	Admixed	9	—	Yes
	10	*C. monticola* (Ijen)	1.0	1.0	1.0	1.0	—	10 + 11 + 12	—	Yes
	11	*C. monticola* (Salak)	1.0	1.0	1.0	1.0	Admixed	10 + 11 + 12	—	Yes
	12	*C. monticola* (Slamet)	1.0	1.0	1.0	1.0	—	10 + 11 + 12	—	Yes
	13	*C.* sp. nov. 1 (Gede)	1.0	1.0	1.0	1.0	—	13	Yes	Yes
	14	*C. maxi* (Ijen)	1.0	1.0	1.0	1.0	na	na	na	Yes
	15	*C. maxi* (Lesser Sundas)	1.0	1.0	1.0	1.0	na	na	na	No
Sumatra	16	*C. beccarii* (Singgalang)	1.0	1.0	1.0	1.0	—	16	Yes	No
	17	*C. beccarii* (Tujuh)	1.0	1.0	1.0	1.0	—	17	Yes	Yes
	18	*C. hutanis* (Leuser)	1.0	1.0	1.0	1.0	—	18	Yes	Yes
	19	*C. lepidura* (Tujuh)	1.0	1.0	1.0	1.0	—	19	Yes	Yes
	20	*C. vosmaeri* (Bangka)	1.0	1.0	1.0	1.0	—	20	Yes	Yes
	21	*C. neglecta* (Tujuh)	1.0	1.0	1.0	1.0	na	na	na	Yes
	22	*C.* cf. *neglecta* (Borneo)	1.0	1.0	1.0	1.0	na	na	na	Yes
	23	*C. paradoxura* (Singgalang)	1.0	1.0	1.0	1.0	—	23	Yes	Yes
	24	*C. paradoxura* (Tujuh)	1.0	1.0	1.0	1.0	—	24	Yes	Yes
	25	*C.* sp. nov. 2 (Singgalang)	1.0	1.0	1.0	1.0	—	25	Yes	Yes

$P < 0.001$; 0.13 migrants per generation), but was not significant from *C. monticola* into *C.* sp. nov. 1 (Table 2; LLR = 0, n.s.). The model of no migration was the least supported model. These results suggest that violations of the assumption of no gene flow among BPP delimited taxa (Yang & Rannala 2014) are minimal because BPP will infer one species when the migration rate is very high (e.g. ≫1 immigrant per generation), while moderate amounts of immigration (e.g. ≪1 immigrant per generation) had little impact on BPP performance (Zhang *et al.* 2011).

Discussion

Our documentation of previously unrecognized shrew diversity and relationships on Java and Sumatra demonstrates substantial within-island endemism and diversification in a continental island system. These results, combined with those of other studies (e.g. Gorog *et al.* 2004; Esselstyn *et al.* 2013), contradict a priori expectations for low levels of interisland diversification and large species ranges. Our phylogenetic inferences suggest at least five intra-island speciation events in *Crocidura* on Java and Sumatra. If we take our BPP results literally, they support up to eight more intra-island speciation events. These inferences suggest the existence of isolating mechanisms that operate within islands and produce species. As we argue above, if barriers within islands are sufficient to generate speciation, then between-island mechanisms also should be sufficient. Therefore, past interpretations of faunal differences between islands of the Sunda shelf (e.g. Heaney

Table 2 Mean number of migrants per generation between geographically sympatric *Crocidura* populations using IMA2

From population	To population	Migrants per generation
C. brunnea (Slamet)	*C. orientalis* (Slamet)	[†]0.2069 (0.08–0.45)
C. orientalis (Slamet)	*C. brunnea* (Slamet)	0.0007 (0.00–0.12)
C. monticola (Gede)	*C. sp. nov. 1* (Gede)	0.0003 (0.00–0.08)
C. sp. nov. 1 (Gede)	*C. monticola* (Gede)	[†]0.1290 (0.05–0.27)

Results are based on eight nuclear loci. The 95% highest posterior density is shown in parentheses.
[†]Migration rates that are significantly different from zero at the *P* < 0.001 level in LLR tests (Nielsen and Wakeley 2001).

1986; Okie & Brown 2009) may have overemphasized the importance of local extinction of wide-range species while discounting the importance of intra- and interisland diversification. If between-island speciation were rampant, it would generate the same patterns (a species is present on one island, but not another) that have been interpreted as extinction of local populations. Although the fossil record clearly demonstrates that extinction has happened on Sundaic islands (Cranbrook 2010), we suggest that speciation is also an important factor contributing to patterns of β-diversity.

Knowledge of the timing and extent of possible inter- and intra-island barriers has been greatly expanded in the last few years. For instance, paleoecological data support the presence of continuous lowland dipterocarp rainforest between Sumatra and Borneo, but not Java, when the Sunda Shelf was exposed during glacial maxima (Raes *et al.* 2014). That study contradicted an earlier supposition that extensive savannah habitats isolated both Sumatra and Java from Borneo during glacial periods (Heaney 1991). Rather, data from Raes *et al.* (2014) are consistent with recurrent interisland barriers to forest-dependent species between Java and other Sundaic regions, but not between Sumatra and Borneo. Also, new tectonic reconstructions support recurrent isolation of Javan and Sumatran montane blocks as a result of cyclical marine inundation (above present sea levels) and volcanic activity up to the Pliocene–Pleistocene boundary (Hall 2009; de Bruyn *et al.* 2014), providing a possible mechanism for past isolation within modern islands. Finally, reconstruction of Sundaland rainforest coverage at the LGM suggests persistent and extensive forest coverage in Borneo, but highly diminished and fragmented forest coverage in Java, and an intermediate level of forest contraction in Sumatra (Cannon *et al.* 2009; de Bruyn *et al.* 2014). These processes may have isolated forest fragments both between and within islands to varying extents, thereby producing idiosyncratic patterns of genetic diversity in extant lineages (Sheldon *et al.* 2015).

Distinct patterns of micro-endemism are apparent at the species and population levels on both Java and Sumatra. We recovered evidence for species-level divergence between populations of *Crocidura paradoxura* and *C. beccarii* from Mts. Singgalang and Tujuh. These peaks are separated by ~190 km and are connected by continuous montane or lower montane forest. Although the divergences between populations of the two shrew species may simply reflect isolation by distance, they potentially represent divergence between closely related allopatric species. However, isolation on mountains separated by a matrix of lowland forest or nonforest habitats does not necessarily generate such levels of genetic divergence. For example, analyses of shrew and rodent populations distributed among disjunct montane forests in Kenya, at similar distances between mountains, did not support independent evolutionary lineages (Demos *et al.* 2014a,b, 2015).

Environmentally, Java is very different from Sumatra. Javan mountains are volcanic and typically separated by wide expanses of drier lowland habitats, while Sumatran peaks are better connected by mountainous habitats, providing a more obvious potential explanation for genetic divergence between Javan populations. Thus, although we anticipated detecting distinct genetic populations on isolated Javan peaks, the divergence observed between Sumatran populations was more surprising. Nevertheless, the phylogenetic relationships reflecting within-island speciation and the small species ranges we find on both islands are not consistent with the expectations of large species ranges and limited isolation in the supposedly well-connected Sundaic system.

Diversity in Sundaland *Crocidura*

Through a combination of improved sampling from new fieldwork and multilocus molecular analyses, we found that shrew species diversity on the South-East Asian continental shelf islands of Java and Sumatra is underestimated, indicating the need for additional biodiversity surveys and taxonomic revisions. Estimates of *Crocidura* diversity based on the most recent comprehensive morphological revision (Ruedi 1995) reported only three *Crocidura* species on Java, two of which were proposed as endemic. Our study, which expanded the data set from Esselstyn *et al.* (2013), identified at least six species on Java (*C. abscondita*, *C. brunnea*, *C. maxi*, *C. monticola*, *C. orientalis* and an undescribed species), including five endemic species. Using coalescent analyses in BPP, we delimited an additional six lineages on Java that may represent distinct species (three lineages in *C. brunnea*, four in *C. monticola* and two in *C. orientalis*). Thus, at least six species are present on Java, but as many as 12 may be represented in our current

sampling. We came to a similar conclusion for Sumatra, where Ruedi (1995) diagnosed six species, five of which he considered endemic. Our results recovered at least seven species on Sumatra and a small neighbouring island (*C. beccarii*, *C. hutanis*, *C. lepidura*, *C. neglecta*, *C. paradoxura*, *C. vosmaeri* and an undescribed species). All of these species are single-island endemics. Our BPP analyses also delimited two lineages each of *C. paradoxura* and *C. beccarii*, which suggests that nine or more species of *Crocidura* may be endemic to Sumatra. The large number of lineages delimited by BPP analyses with a geographically sparse set of samples emphasizes the need for additional specimen collection.

Our expanded geographic sampling that included specimens from the eastern extremity of Java (Mt. Ijen) also made clear the status of *Crocidura maxi*. This species was previously recognized from East Java and the Lesser Sunda Islands (Kitchener *et al.* 1994). Esselstyn *et al.* (2013, 2014), however, identified specimens from Mt. Gede (the first West Javan record) as *C. maxi* and reported that they were not closely related to animals from the Lesser Sundas. In this study, we obtained new specimens of *C. maxi* from Mt. Ijen, which are closely related to the Lesser Sunda shrews, but not the Mt. Gede series. This clarifies that *C. maxi* is indeed present in East Java and the Lesser Sundas, as Kitchener *et al.* (1994) indicated, while the Mt. Gede series from Esselstyn *et al.* (2013) is a new species (*C. sp. nov. 1*).

Syntopic sister species

On Mt. Gede, Java, *Crocidura monticola* was collected together with a genetically and phenotypically distinguishable (see figs 3 and 4 in Esselstyn *et al.* 2014), but as yet undescribed species, *C. sp. nov. 1*. We found an apparent pattern of partial elevational overlap between these species, with 16 *C. monticola* sampled at 1377 m, 32 *C. monticola* and 22 *C. sp. nov. 1* sampled at 1611 m, and 13 *C. sp. nov. 1* sampled at 1950 m. These two species have an uncorrected pairwise *cyt-b* distance of 3.8%. Median fossil calibrated multilocus divergence estimates from Esselstyn *et al.* (2013) ranged from 178 000 to 515 000 generations ago, placing divergence in the Pleistocene. Where the two species are syntopic, our results from IMA2 analyses support moderate gene flow from *C. sp. nov. 1* into *C. monticola* and a very low level of gene flow in the other direction (Table 2). A migration rate LLR test in IMA2 was only significant for gene flow from *C. sp. nov. 1* into *C. monticola*. These results suggest that population divergence may have occurred with gene flow, possibly along a single elevational gradient. It is surprising that these two species evolved diagnosable morphological differences since their very recent divergence (Esselstyn *et al.* 2014). The fact that morphological disparity has evolved in such a short time in an otherwise morphologically conservative group suggests that selection is involved. We suggest this species pair represents a plausible example of ecological speciation (e.g. Nosil 2012) that could be tested with more data.

Conclusions

Our phylogenetic and phylogeographic analyses found high levels of previously unrecognized inter- and intra-island diversity. Inferences from multiple analyses strongly support at least seven Sumatran and six Javan *Crocidura* lineages as valid species. All but one of these species is endemic to a single island, and several species are only known from a single mountain. Two pairs of sister taxa on each of these islands suggest that at least five within-island speciation events have occurred. The most recent of these events generated two morphologically distinct species, with current populations occurring syntopically in at least one site. The newly recognized patterns of endemism, in which no species of *Crocidura* is widespread on the Sunda shelf, indicate that evolutionary processes on these islands may be more similar to those reported for oceanic archipelagos (Heaney *et al.* 2011; Justiniano *et al.* 2015), where species ranges are often smaller than the islands themselves. This is in stark contrast to the traditional expectation that species should be widespread in continental island systems such as Sundaland, and warrants reconsideration of speciation as part of the processes that generated β-diversity across this island system.

Acknowledgments

References

Achmadi AS, Maryanto I, Maharadatunkamsi (2012) Systematic and descriptions of new species within genus *Maxomys* from East Kalimantan, Borneo Island. *Treubia*, **39**, 1–28.

Bland LM, Collen B, Orme CDL, Bielb J (2015) Predicting the conservation status of data-deficient species. *Conservation Biology*, **29**, 250–259.

Brito D (2010) Overcoming the Linnean shortfall: data deficiency and biological survey priorities. *Basic and Applied Ecology*, **11**, 709–713.

Brown JH (1986) Two decades of interaction between the MacArthur-Wilson model and the complexities of mammalian distributions. *Biological Journal of the Linnean Society*, **28**, 231–251.

de Bruyn M, Stelbrink B, Morley RJ *et al.* (2014) Borneo and Indochina are major evolutionary hotspots for Southeast Asian biodiversity. *Systematic Biology*, **63**, 879–901.

Burbrink FT, Yao H, Ingrasci M *et al.* (2011) Speciation at the Mogollon Rim in the Arizona Mountain Kingsnake (*Lampropeltis pyromelana*). *Molecular Phylogenetics and Evolution*, **60**, 445–454.

Burnham KP, Anderson DR (2002) *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach.* Springer-Verlag, New York.

Camargo A, Morando M, Avila LJ, Sites JW (2012) Species delimitation with ABC and other coalescent-based methods: a test of accuracy with simulations and an empirical example with lizards of the *Liolaemus darwinii* complex (Squamata: Liolaemidae). *Evolution*, **66**, 2834–2849.

Cannon CH, Morley RJ, Bush ABG (2009) The current refugial rainforests of Sundaland are unrepresentative of their biogeographic past and highly vulnerable to disturbance. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 11188–11193.

Carstens BC, Stoute HN, Reid NM (2009) An information-theoretical approach to phylogeography. *Molecular Ecology*, **18**, 4270–4282.

Corbet GB, Hill JE (1992) *The Mammals of the Indomalayan Region: A Systematic Review.* Oxford University Press, Oxford.

Cranbrook E (2010) Later quaternary turnover of mammals in Borneo: the zooarchaeological record. *Biodiversity Conservation*, **19**, 373–391.

Darriba D, Taboada GL, Doallo R, Posada D (2012) jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods*, **9**, 772.

Demos TC, Agwanda B, Hickerson MJ (2014a) Integrative taxonomy of *Hylomyscus* (Rodentia: Muridae) and description of a new species from western Kenya. *Journal of Mammalogy*, **95**, E1–E15.

Demos TC, Kerbis Peterhans JC, Agwanda B, Hickerson MJ (2014b) Uncovering cryptic diversity and refugial persistence among small mammal lineages across the Eastern Afromontane biodiversity hotspot. *Molecular Phylogenetics and Evolution*, **71**, 41–54.

Demos TC, Kerbis Peterhans JC, Joseph TA *et al.* (2015) Comparative population genomics of African montane forest mammals support population persistence across a climatic gradient and Quaternary climatic cycles. *PLoS ONE*, **10**, e0131800.

Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, **29**, 1969–19673.

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

Esselstyn JA, Goodman SM (2010) New species of shrew (Soricidae: *Crocidura*) from Sibuyan Island, Philippines. *Journal of Mammalogy*, **91**, 1467–1472.

Esselstyn JA, Timm RM, Brown RM (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 JA, Oliveros CH, Moyle RG *et al.* (2010) Integrating phylogenetic and taxonomic evidence illuminates complex biogeographic patterns along Huxley's modification of Wallace's Line. *Journal of Biogeography*, **37**, 2054–2066.

Esselstyn JA, Achmadi AS, Rowe KC (2012) Evolutionary novelty in a rat with no molars. *Biology Letters*, **8**, 990–993.

Esselstyn JA, Maharadatunkamsi, Achmadi AS *et al.* (2013) Carving out turf in a biodiversity hotspot: multiple, previously unrecognized shrew species co-occur on Java Island, Indonesia. *Molecular Ecology*, **22**, 4972–4987.

Esselstyn JA, Achmadi AS, Maharadatunkamsi (2014) A new species of shrew (Soricomorpha: *Crocidura*) from West Java, Indonesia. *Journal of Mammalogy*, **95**, 216–224.

Esselstyn JA, Achmadi AS, Handika H, Rowe KC (2015) A hog-nosed shrew rat (Rodentia: Muridae) from Sulawesi Island, Indonesia. *Journal of Mammalogy*, **96**, 895–907.

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.

Flot JF (2010) SEQPHASE: a web tool for interconverting phase input/output files and fasta sequence alignments. *Molecular Ecology Resources*, **10**, 162–166.

Gao H, Bryc K, Bustamante CD (2011) On identifying the optimal number of population clusters via the deviance information criterion. *PLoS ONE*, **6**, e21014.

Garrick RC, Sunnucks P, Dyer RJ (2010) Nuclear gene phylogeography using PHASE: dealing with unresolved genotypes, lost alleles, and systematic bias in parameter estimation. *BMC Evolutionary Biology*, **10**, 118.

Giarla TC, Voss RS, Jansa SA (2014) Hidden diversity in the Andes: comparison of species delimitation methods in montane marsupials. *Molecular Phylogenetics and Evolution*, **70**, 137–151.

Gorog AJ, Sinaga MH, Engstrom MD (2004) Vicariance or dispersal? Historical biogeography of three Sunda shelf marine rodents (*Maxomys surifer*, *Leopoldamys sabanus* and *Maxomys whiteheadi*). *Biological Journal of the Linnean Society*, **81**, 91–109.

Grant PR, Grant BR (2011) *How and Why Species Multiply: The Radiation of Darwin's Finches.* Princeton University Press, Princeton, New Jersey.

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

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

Heaney LR (1991) A synopsis of climate and vegetational changes in Southeast Asia. *Climatic Change*, **19**, 53–61.

Heaney LR (2000) Dynamic disequilibrium: a long-term, large-scale perspective on the equilibrium model of island biogeography. *Global Ecology and Biogeography*, **9**, 59–74.

Heaney LR (2007) Is a new paradigm emerging for oceanic island biogeography? *Journal of Biogeography*, **34**, 753–757.

Heaney LR, Balete DS, Rickart EA *et al.* (2011) Seven new species and a new subgenus of forest mice (Rodentia: Muridae: *Apomys*) from Luzon Island. *Fieldiana Life and Earth Sciences*, **2**, 1–60.

Hey J (2010) Isolation with migration models for more than two populations. *Molecular Biological Evolution*, **27**, 905–920.

Hey J, Nielsen R (2007) Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 2785–2790.

Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources*, **9**, 1322–1332.

IUCN (2015) *The IUCN Red List of Threatened Species*. Version 2015-4. Available from http://www.iucnredlist.org.

Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, **23**, 1801–1806.

Jenkins PD (1982) A discussion of Malayan and Indonesian shrews of the genus *Crocidura* (Insectivora: Soricidae). *Zoologische Mededelingen*, **56**, 267–279.

Jenkins PD, Lunde DP, Moncrieff CB (2009) Descriptions of new species of *Crocidura* (Soricomorpha: Soricidae) from mainland Southeast Asia, with synopses of previously described species and remarks on biogeography. *Bulletin of the American Museum of Natural History*, **331**, 356–405.

Jenkins PD, Abramov AV, Rozhnov VV, Olsson A (2010) A new species of *Crocidura* (Soricomorpha: Soricidae) from southern Vietnam and north-eastern Cambodia. *Zootaxa*, **2345**, 60–68.

Jenkins PD, Abramov AV, Bannikova AA, Rozhnov VV (2013) Bones and genes: resolution problems in three Vietnamese species of *Crocidura* (Mammalia, Soricomorpha, Soricidae) and the description of an additional new species. *ZooKeys*, **313**, 61–79.

Justiniano R, Schenk JJ, Balete DS *et al.* (2015) Testing diversification models of endemic Philippine forest mice (*Apomys*) with nuclear phylogenies across elevational gradients reveals repeated colonization of isolated mountain ranges. *Journal of Biogeography*, **42**, 51–64.

Kerhoulas NJ, Gunderson AM, Olson LE (2015) Complex history of isolation and gene flow in hoary, Olympic, and endangered Vancouver Island marmots. *Journal of Mammalogy*, **96**, 810–826.

Kitchener DJ, Hisheh S, Schmitt LH, Suyanto A (1994) Shrews (Soricidae: *Crocidura*) from the Lesser Sunda Islands, and South-East Maluku, Eastern Indonesia. *Australian Mammalogy*, **17**, 7–17.

Kopelman NM, Mayzel J, Jakobsson M *et al.* (2015) CLUMPAK: a program for identifying clustering modes and packaging population structure inferences across *K*. *Molecular Ecology Resources*, **15**, 1179–1191.

Kumar S, Subramanian S (2002) Mutation rates in mammalian genomes. *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 803–808.

Lanfear R, Calcott B, Ho SYW, Guindon S (2012) PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution*, **29**, 1695–1701.

Leaché AD, Fujita MK (2010) Bayesian species delimitation in West African forest geckos (*Hemidactylus fasciatus*). *Proceedings of the Royal Society B-Biological Sciences*, **277**, 3071–3077.

Leonard JA, den Tex R-J, Hawkins MTR *et al.* (2015) Phylogeography of vertebrates on the Sunda Shelf: a multi-species comparison. *Journal of Biogeography*, **42**, 871–879.

MacArthur RH, Wilson EO (1967) *The Theory of Island Biogeography*. Princeton University Press, Princeton, New Jersey.

Meirmans PG (2015) Seven common mistakes in population genetics and how to avoid them. *Molecular Ecology*, **13**, 3223–3231.

Merckx VSFT, Hendriks KP, Beentjes KK *et al.* (2015) Evolution of endemism on a young tropical mountain. *Nature*, **524**, 347–350.

Nielsen R, Wakeley J (2001) Distinguishing migration from isolation: a Markov chain Monte Carlo approach. *Genetics*, **158**, 885–896.

Nosil P (2012) *Ecological Speciation*. Oxford University Press, New York.

Okie JG, Brown JH (2009) Niches, body sizes, and the disassembly of mammal communities on the Sunda Shelf islands. *Proceedings of the National Academy of Sciences of the United States of America*, **106**(Suppl.), 19679–19684.

Oliveros CH, Moyle RG (2010) Origin and diversification of Philippine bulbuls. *Molecular Phylogenetics and Evolution*, **54**, 822–832.

Omar H, Hashim R, Bhassu S, Ruedi M (2013) Morphological and genetic relationships of the *Crocidura monticola* species complex (Soricidae: Crocidurinae) in Sundaland. *Mammalian Biology*, **78**, 446–454.

Papadopoulou A, Knowles LL (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. *Nature*, **69**, 1501–1517.

Piper PJ, Ochoa J, Lewis H, Paz V, Ronquillo WP (2008) The first evidence for the past presence of the tiger *Panthera tigris* (L.) on the island of Palawan, Philippines: extinction in an island population. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **264**, 123–127.

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

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

Raes N, Cannon CH, Hijmans RJ *et al.* (2014) Historical distribution of Sundaland's Dipterocarp rainforests at Quaternary glacial maxima. *Proceedings of the National Academy of Sciences of the United States of America*, **111**, 16790–16795.

Rambaut A, Suchard MA, Xie D, Drummond AJ (2014) Tracer v1.6, Available from http://beast.bio.ed.ac.uk/Tracer.

Roberts TE, Lanier HC, Sargis EJ, Olson LE (2011) Molecular phylogeny of treeshrews (Mammalia: Scandentia) and the timescale of diversification in Southeast Asia. *Molecular Phylogenetics and Evolution*, **60**, 358–372.

Ronquist F, Teslenko M, van der Mark P *et al.* (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, **61**, 539–542.

Rosenberg NA (2004) DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology Notes*, **4**, 137–138.

Rosenzweig ML (1995) *Species Diversity in Space and Time*. Cambridge University Press, Cambridge.

Rowe KC, Achmadi AS, Esselstyn JA (2016) A new genus and species of omnivorous rodent (Muridae: Murinae) from Sulawesi, nested within a clade of endemic carnivores. *Journal of Mammaolgy*, **97**, 978–991.

Ruedi M (1995) Taxonomic revision of shrews of the genus *Crocidura* from the Sunda Shelf and Sulawesi with a description of two new species (Mammalia: Soricidae). *Zoological Journal of the Linnean Society*, **115**, 211–265.

Ruedi M (1996) Phylogenetic evolution and biogeography of Southeast Asian shrews (genus *Crocidura*: Soricidae). *Biological Journal of the Linnean Society*, **58**, 197–219.

Ruedi M, Fumagalli L (1996) Genetic structure of Gymnures (genus *Hylomys*; Erinaceidae) on continental islands of Southeast Asia: historical effects of fragmentation. *Journal of Zoological Systematics and Evolutionary Research*, **34**, 153–162.

Schluter D (2000) *The Ecology of Adaptive Radiation*. Oxford University Press, Oxford.

Sheldon FH, Lim HC, Moyle RG (2015) Return to the Malay Archipelago: the biogeography of Sundaic rainforest birds. *Journal of Ornithology*, **156**, S91–S113.

Sheth SN, Lohmann LG, Distler T, Jiménez I (2012) Understanding bias in geographic range size estimates. *Global Ecology and Biogeography*, **21**, 732–742.

Stelbrink B, Albrect C, Hall R, von Ritelen 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.

Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics*, **68**, 978–989.

Sukumaran J, Holder MT (2010) DendroPy: a Python library for phylogenetic computing. *Bioinformatics*, **26**, 1569–1571.

den Tex R-J, Leonard JA (2013) A molecular phylogeny of Asian barbets: speciation and extinction in the tropics. *Molecular Phylogenetics and Evolution*, **68**, 1–13.

Wallace AR (1876) *The Geographical Distribution of Animals, With a Study of the Relations of Living and Extinct Faunas as Elucidating the Past Changes of the Earth's Surface*. Harper & Brothers, New York.

Wallace AR (1881) *Island Life*. Harper & Brothers, New York.

Waples RS, Gaggiotti O (2006) What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Molecular Ecology*, **15**, 1419–1439.

Whittaker RJ, Fernández-Palacios JM (2007) *Island Biogeography: Ecology, Evolution and Conservation*. Oxford University Press, Oxford.

Wilting A, Sollmann R, Meijaard E *et al.* (2012) Mentawai's endemic, relictual fauna: is it evidence for Pleistocene extinctions on Sumatra? *Journal of Biogeography*, **39**, 1608–1620.

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

Yang Z, Rannala B (2014) Unguided species delimitation using DNA sequence data from multiple loci. *Molecular Biology and Evolution*, **31**, 3125–3135.

Zhang C, Zhang D-X, Zhu T, Yang Z (2011) Evaluation of a Bayesian coalescent method of species delimitation. *Systematic Biology*, **60**, 747–761.

Zwickl DJ (2006) *Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion*. PhD Dissertation, The University of Texas at Austin, Austin, Texas.

Data accessibility

T.C.D., T.C.G. and J.A.E. designed the study; J.A.E., A.S.A., H.H., M. and K.C.R. conducted fieldwork and identified specimens; T.C.D., T.C.G. and J.A.E. sequenced DNA; T.C.D. and T.C.G. analysed the data; T.C.D., J.A.E. and K.C.R. wrote the manuscript with editorial contributions from T.C.G., A.S.A. and H.H.

Supporting information

Additional supporting information may be found in the online version of this article.

Appendix S1 List of the museum voucher numbers, localities, elevations, and GenBank accession numbers for all specimens used in this study. NA indicates samples intentionally not included in the study and blank cells indicate failure of polymerase chain reaction amplification.

Table S1 Results of pairwise IMA2 L-mode analyses using ranked nested-models of migration for *Crocidura brunnea* and *C. orientalis*.

Table S2 Results of pairwise IMA2 L-mode analyses using ranked nested-models of migration for *Crocidura monticola* and *C. sp. nov. 1*.

Fig. S1 Maximum likelihood gene tree estimates of phased alleles from Southeast Asian shrews (genus *Crocidura*) for (A) *ApoB*, (B) *BDNF*, (C) *BRCA1*, (D) *GHR10*, (E) *MCGF*, (F) *PTGER4*, (G) *RAG1*, and (H) *vWF*.

Fig. S2 Bayesian gene tree estimates of phased alleles from Southeast Asian shrews (genus *Crocidura*) for (A) *ApoB*, (B) *BDNF*, (C) *BRCA1*, (D) *GHR10*, (E) *MCGF*, (F) *PTGER4*, (G) *RAG1*, and (H) *vWF*.

Fig. S3 Results from STRUCTURE analyses of eight nuclear loci from 12 *Crocidura* species (A–G) with the number of populations (*K*) varying from 1–7.