Zoological Journal of the Linnean Society, 2022, **194**, 478–501. With 7 figures.

Evolutionary history of Sundaland shrews (Eulipotyphla: Soricidae: *Crocidura*) with a focus on Borneo

ARLO HINCKLEY[1,*], MIGUEL CAMACHO-SANCHEZ[1,2], MANUEL RUEDI[3], MELISSA T. R. HAWKINS[4], MADELEINE MULLON[5], ANNA CORNELLAS[1], FRED TUH YIT YUH[6] and JENNIFER A. LEONARD[1,*]

[1]*Conservation and Evolutionary Genetics Group, Estación Biológica de Doñana (EBD-CSIC), Seville, Spain*
[2]*Instituto Andaluz de Investigación y Formación Agraria, Pesquera, Alimentaria y de la Producción Ecológica (IFAPA) Centro Las Torres, Alcalá del Río, Spain*
[3]*Natural History Museum of Geneva, Switzerland*
[4]*National Museum of Natural History, Department of Vertebrate Zoology, Smithsonian Institution, USA*
[5]*Harvard Museum of Comparative Zoology, USA*
[6]*Sabah Parks, Kota Kinabalu, Sabah, Malaysia*

Received 2 December 2020; revised 18 May 2021; accepted for publication 15 June 2021

The hyperdiverse shrew genus *Crocidura* is one of few small mammal genera distributed across Sundaland and all of its boundaries. This represents a rare opportunity to study the geological history of this region through the evolutionary history of these shrews. We generate a phylogeny of all recognized species of Sundaland *Crocidura* and show that most speciation events took place during the Pleistocene, prior to the inundation of the Sunda Shelf around 400 000 years ago. We find east–west differentiation within two separate lineages on Borneo, and that the current taxonomy of its two endemic species does not reflect evolutionary history, but ecophenotypic variation of plastic traits related to elevation. Sulawesi shrews are monophyletic, with a single notable exception: the black-footed shrew (*C. nigripes*). We show that the black-footed shrew diverged from its relatives on Borneo recently, suggesting a human-assisted breach of Wallace's line. Overall, the number of *Crocidura* species, especially on Borneo, probably remains an underestimate.

ADDITIONAL KEYWORDS: biogeography – cryptic species – divergence – phylogenetic systematics – South-East Asia – species delineation – speciation.

INTRODUCTION

The biogeographic region of Sundaland (Borneo, Sumatra, the Malay Peninsula, Java, Palawan and associated islands) is important to our understanding of evolution. Sundaland inspired Alfred Russel Wallace to independently arrive at the theory of evolution by natural selection, and has been instrumental in the foundation of modern biogeography (Wallace, 1869). This region is defined by three biogeographic limits. Wallace's Line is a sharp biogeographical break that divides Sulawesi and the Lesser Sundas from Sundaland (Wallace, 1869). Huxley's line separates Sundaland and the oceanic islands of the Philippines, and the Isthmus of Kra divides Sundaland from Indochina (Fig. 1). The first two are oceanic biogeographic limits and the latter is a terrestrial transition, possibly shaped by rainfall variation and associated habitat differences (Woodruff & Turner, 2009; Lim *et al.*, 2020).

The geological structure of Sundaland, as a series of currently isolated land masses that were previously contiguous, makes the region excellent for testing alternative hypotheses that explain the observed high biodiversity in the tropics (Hall, 2013; Sheldon *et al.*, 2015). The continental Sunda Shelf was repeatedly inundated when polar ice-sheets melted during warm Pleistocene periods, with the resulting rise of sea levels isolating the land masses periodically since about 400 thousand years ago (kya) (Husson *et al.*, 2020). Interestingly, recent studies on mammals

*Corresponding authors. E-mail: arlo@ebd.csic.es; JLeonard@ebd.csic.es

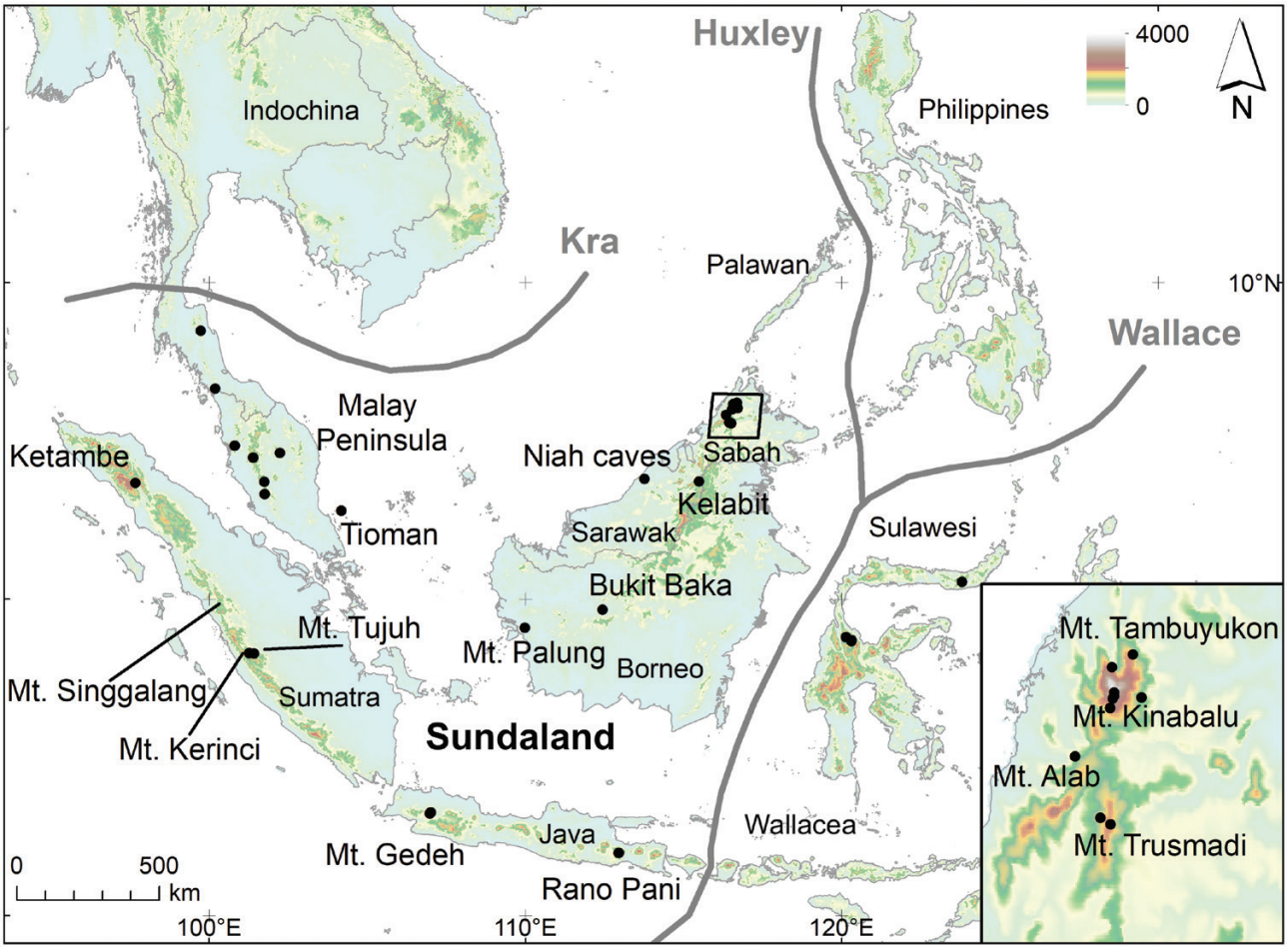

Figure 1. Map of tropical east Asia indicating biogeographic regions and sample localities in this study.

have found that populations on different landmasses diverged during the Miocene and Pliocene, long before the shelf was first inundated, and so are too ancient to be explained by the Pleistocene pump hypothesis (Den Tex *et al.*, 2010; Roberts *et al.*, 2011; Mason *et al.*, 2019; Hinckley *et al.*, 2020). Ecological isolation driven by open terrestrial habitats, such as grassland, swamp and/ or savanna, might have predated physical isolation by marine barriers (Cannon *et al.*, 2009; Slik *et al.*, 2011). Alternatively, ancient divergences could derive from rainforest refugia on or around mountains through the cooler and drier Oligocene and Pliocene (Sheldon *et al.*, 2015). If isolation of these rainforest pockets was a main driver of diversification in Sundaland, it could also manifest between refugia within landmasses. The impact of these ancient conditions will impact species differently, depending on their habitat requirements and dispersal capabilities (Mason *et al.*, 2019; Cros *et al.*, 2020; Lim *et al.*, 2020). Forest specialists may have been subject to more population fragmentation with longer periods of low effective population sizes (ESS), increasing the intensity of genetic drift (Després, 2019; Mason *et al.*, 2019). Further, island populations experience greater rates of extinction relative to continental ones, even for small vertebrates in Sundaland (Heaney, 1984; Den Tex *et al.*, 2010) so colonization or re-colonization following local extinction could also have had an important impact on biogeographical patterns (Leonard *et al.*, 2015). Incomplete taxonomy may obscure underlying patterns. This shortfall is more acute for smaller organisms, many of which have narrow niches or restricted ranges (Riddle *et al.*, 2011).

An understanding of the species diversity is an important first step to a deeper understanding of the evolution of biodiversity in this region. Unlike other vertebrates, mammal species are generally distinguished by subtle morphological characteristics that hinder their proper taxonomic classification (Patterson, 2002). A particularly morphologically conserved group are shrews of the genus *Crocidura* Wagler, 1832. This species-rich genus includes some

widespread species and some narrow endemics (Burgin *et al.*, 2018a). A combination of sampling from new fieldwork and integrative taxonomy has recently increased the number of species in this genus in tropical east Asia, both through splitting widespread species into multiple geographically restricted ones and the discovery of *de novo* species (Ruedi, 1995; Jenkins *et al.*, 2007; Abramov *et al.*, 2008; Jenkins *et al.*, 2009; Esselstyn & Goodman, 2010; Jenkins *et al.*, 2013; Esselstyn & Achmadi, 2014; Demos *et al.*, 2017; Hutterer *et al.*, 2018; Esselstyn *et al.*, 2019; Chen *et al.*, 2020; Yang *et al.*, 2020). This progress is not evenly distributed across the region, and Borneo has been notably absent in recent revisions.

Recently recognized or putative *Crocidura* species are generally restricted to a single habitat on a single landmass in Sundaland (Esselstyn *et al.*, 2013; Demos *et al.*, 2016). Likewise, surprisingly high levels of divergence were found among two Sumatran peaks separated by ~190 km and connected by continuous montane or lower montane forest, which suggests an important role of within-island diversification (Demos *et al.*, 2016). This pattern of high divergence between highland taxa has also been found in other small mammals on Sumatra (Den Tex *et al.*, 2010; Achmadi *et al.*, 2013; Camacho-Sanchez & Leonard, 2020; Hinckley *et al.*, 2020) and Java (Demos *et al.*, 2016). In contrast to species-rich Sumatra and Java, with up to ten and 12 putative species of *Crocidura*, respectively (Demos *et al.*, 2016), only four and three species are recognized on the Malay Peninsula and Borneo (Ruedi, 1995). However, these landmasses have been less sampled than Sumatra and Java (Ruedi, 1995; Omar *et al.*, 2013). In light of Demos *et al.*'s (2016) findings, it is possible that species diversity in Peninsular Malaysia and Borneo are also underestimated.

Crocidura is also interesting from a biogeographic perspective, since it is one of few small mammal genera distributed across the biogeographic breaks of Wallace's and Huxley's Lines, and the Isthmus of Kra, representing a rare opportunity to simultaneously compare the permeability of all three Sundaland boundaries. All *Crocidura* shrews from Sulawesi are endemic and diversified in a single radiation, except the black-footed shrew (*C. nigripes* Miller & Hollister, 1921), thought to be a more recent colonizer (Ruedi *et al.*, 1998; Esselstyn *et al.*, 2019). However, an increased gene sampling and geographic coverage across Borneo is needed to elucidate if its ancestor colonized Sulawesi from the Philippines or Borneo (Demos *et al.*, 2016). Huxley's line seems to have been crossed a single time by *Crocidura* when they colonized the Philippines. Such colonization is thought to have been mediated through Palawan, which served as a biogeographic bridge from the Sunda Shelf islands to the oceanic Philippine islands (Esselstyn *et al.*, 2010; Giarla & Esselstyn, 2015). The number of times *Crocidura* has crossed the Isthmus of Kra has not been addressed. However, a recent phylogeny could suggest multiple colonization events (Chen *et al.*, 2020) from Indochina to Sundaland. Due to sparse geographic datasets and lack of appropriate genetic markers, the timing of the crossings across biogeographic limits of Sundaland are poorly understood.

In this study, we use molecular genetic data from all recognized species of *Crocidura* of Sundaland, Sulawesi and most of the Philippines, to construct a well-resolved dated phylogeny of the genus in this region. These results are placed into geographical context in order to compare the diversity and diversification of white-toothed shrews on the different major land masses and between them, with an emphasis on the species of Borneo. Morphological data of the Bornean endemics *C. foetida* Peters, 1870 and *C. baluensis* Thomas, 1898 is integrated with genetic data in order to gain insight into the evolution of this lineage and assess the effect of elevation on phenotypic divergence.

Study system

Despite its old age, large size and geographic complexity (Sheldon *et al.*, 2015; Whittaker *et al.*, 2017), only three *Crocidura* species are recognized on Borneo: the widespread *C. foetida* and *C. cf. neglecta* Jentink 1888, which are distributed across Borneo's lowlands and hills Sundaland, respectively, and the narrow highland endemic (*C. baluensis*). Species diversity in other vertebrates, including small mammals, is frequently higher on Borneo than Sumatra (Roberts *et al.*, 2011; Achmadi *et al.*, 2013; De Bruyn *et al.*, 2014; Hinckley *et al.*, 2020). The high-elevation Kinabalu shrew (*C. baluensis*) is thought to be restricted to the higher slopes of Mt. Kinabalu (4095 m), and perhaps to other nearby high mountains (Ruedi, 1995). This species shows strong morphological resemblance to Sumatran highland *C. lepidura* Lyon, 1908 (Ruedi, 1995), mirroring the pattern of morphological convergence observed in *Rattus baluensis* Thomas, 1894 (Mt. Kinabalu, Borneo) and *Rattus korinchi* Robinson & Kloss, 1916 (Sumatra) (Musser, 1986). Merckx *et al.*, (2015) showed that Mt. Kinabalu endemics are a mix of immigrant, pre-adapted lineages (eccentric endemics) and descendants from local, lowland ancestors (centric endemics). Recently, Camacho-Sanchez & Leonard (2020) have shown that the origin of the Bornean endemic *R. baluensis* is recent and its genetic diversity is nested within the diversity of local lowland *R. tiomanicus* Miller, 1900. However, other Bornean mountain endemic mammals are highly divergent from their lowland sister-taxa (e.g. *Sundasciurus everetti* Thomas, 1890, Hawkins *et al.*, 2016b; *Tupaia montana* Thomas, 1892 Roberts *et al* 2011).

The Bornean shrew (*C. foetida*) is assumed to be widespread throughout the island, since it has

been recorded in Sabah, Sarawak and east and west Kalimantan from sea level to 1900 m a.s.l. (Burgin *et al.*, 2018a; William-Dee, 2019). Three subspecies are recognized. The nominal form was described from West Kalimantan and is potentially distributed throughout the lowlands of west and central Borneo and perhaps also in the north; *C. f. doriae* was described from an unknown locality in Sarawak and is potentially distributed throughout mountains of north Borneo (up to 1900 m a.s.l.); and *C. f. kelabit*, described from Bario and restricted to the Kelabit Highlands (Ruedi, 1995; Burgin *et al.*, 2018a).

The neglected shrew (*C. cf. neglecta*) from Borneo was supported as a distinct species from *C. neglecta* on Sumatra in species delimitation analyses performed by Demos *et al.* (2016). However, these analyses included only two specimens from Mt. Kinabalu and no intermediate populations between extreme north-east Borneo and Sumatra, which would be desirable to assess if this divergent lineage is distributed across Borneo or if there is isolation by distance (Sukumaran & Knowles, 2017; Mason *et al.*, 2020).

Four species of *Crocidura* are found on Peninsular Malaysia, according to the latest systematic review performed for this landmass (Ruedi, 1995). Two are endemic (*C. malayana* Robinson & Kloss, 1911 and *C. negligens* Robinson & Kloss, 1914); and two are thought to be widespread (*Crocidura cf. neglecta* and *C. fuliginosa* Blyth, 1855). Recent sampling and genetic work suggest that both of these widespread species contain multiple, distinct lineages compatible with multiple species (Bannikova *et al.*, 2011; Omar *et al.*, 2013; Burgin *et al.*, 2018a; Chen *et al.*, 2020). The occurrence of a fifth, long-tailed species in the higher slopes of the Cameron Highlands, Peninsular Malaysia, has generated much taxonomic discussion. Different authors have associated these long-tailed specimens with *C. aequicauda* Robinson & Kloss, 1918 (a junior synonym of *C. paradoxura* Dobson, 1887; Burgin *et al.*, 2018a), *C. attenuata* Milne-Edwards, 1870 (Robinson & Kloss, 1918; Jenkins, 1976, 1982; Davison, 1984) or *C. negligens* (Ruedi, 1995). However, these specimens have never been sequenced, so this population could also represent an undescribed highland endemic.

Demos *et al.* (2016, 2017) identified at least six species on Java: *C. abscondita* Esselstyn *et al.*, 2014, *C. brunnea* Jentink, 1888, *C. maxi* Sody, 1936, *C. monticola* Peters, 1870, *C. orientalis* Weber, 1890 and *C. umbra* Demos *et al.*, 2017, five of which are endemic. Species delimitation coalescent analyses supported six additional lineages on Java that may represent distinct species: three lineages in *C. brunnea*, four in *C. monticola* and two in *C. orientalis*. This represents a remarkable increase from three to nine species since the previous taxonomic review (Ruedi, 1995).

A similar scenario was found on Sumatra, where Ruedi (1995) diagnosed six species, five of which he considered endemic, including one he described (*C. hutanis* Ruedi, 1995). While Demos *et al.* (2016) recovered at least seven endemic species on Sumatra and a small neighbouring island (*C. beccarii* Dobson, 1887, *C. hutanis*, *C. lepidura*, *C. neglecta* Jentink, 1888, *C. paradoxura*, *C. vosmaeri* Jentink, 1888 and an undescribed species, *C. sp. nov. 2*). In addition, species delimitation analyses also suggested additional potential species in the two lineages of *C. paradoxura* and *C. beccarii*, implying an increase of up to three species since Ruedi (1995).

Miller & Hollister (1921) described five species of *Crocidura* on Sulawesi. The next review of this genus, involving extensive fieldwork (Ruedi, 1995), described one new species (*C. musseri*). Subsequently, Esselstyn *et al.*, (2019) described an additional one (*C. caudipilosa*), and suggested two additional undescribed species. These studies brought the final number of putative species of *Crocidura* on Sulawesi to nine: *C. caudipilosa*, 'pale *C. elongata* Miller & Hollister, 1921', 'dark *C. elongata*', 'pale *C. lea* Miller & Hollister, 1921', 'dark *C. lea*', *C. levicula* Miller & Hollister, 1921, *C. musseri*, *C. nigripes* Miller & Hollister, 1921 and *C. rhoditis* Miller & Hollister, 1921. The Lesser Sundas also host two other species east of Wallace's Line: *C. maxi* and *C. tenuis* Muller, 1840. The former is widespread and also present in east Java, while the latter is endemic to Timor. An endemic and possibly extinct species, *C. trichura* Dobson, 1889 has been recorded from Christmas Island, around 350 km south of Java.

Many species of *Crocidura* in the political region of the Philippines were described in the last century: *C. beatus* Miller, 1910, *C. grandis* Miller, 1910, *C. grayi* Dobson, 1890, *C. mindorus* Miller, 1910, *C. negrina* Rabor, 1952, *C. palawanensis* Taylor, 1934. While others were described more recently: *C. ninoyi* Esselstyn & Goodman, 2010 and *C. panayensis* Hutterer, 2007. The Palawan endemic *C. batakorum* Hutterer, 2007 has been shown to be the only species not belonging to the *in situ* Philippine radiation, besides peripherical *C. tanakae* present in Batan and Sabtang, two islands in between Taiwan and Luzon (Esselstyn & Oliveros, 2010). In addition to *C. batakorum*, *C. palawanensis* is also endemic to Palawan, which is biogeographically part of Sundaland. The species *C. grandis* has not been recorded in over a century and its phylogenetic relationships are unknown (Giarla & Esselstyn, 2015).

MATERIAL AND METHODS

Material

Molecular sampling of all recognized extant species of Sundaland and Sulawesi *Crocidura*, and most from

the Lesser Sunda Islands, Philippines and Indochina (Fig. 1), was achieved through a combination of fieldwork, historic museum specimens and sequences deposited in GenBank. Specimens were identified based on their morphology and distributions following Ruedi (1995). Our fieldwork, together with tissue loans from American Museum of Natural History (AMNH), Estación Biológica de Doñana-CSIC (EBD) and A. Gorog, yielded 94 modern samples. Additionally, we obtained 15 tissue samples from historic specimens housed in five different museums: Lee Kong Chian Museum (LKC), Museum of Comparative Zoology of Harvard (MCZ), Museo Nacional de Ciencias Naturales (MNCN), Thailand National History Museum (TNHM) and Museum für Naturkunde (ZMB). In total, 109 samples were included (Supporting Information, Table S1). Data from another 563 samples of 93 species/putative species was downloaded from GenBank for comparison (Supporting Information, Table S1). Samples were collected according to the guidelines of the American Society of Mammalogists (Sikes *et al.*, 2016). Animal care and use committees approved protocols, sampling and field research permits for the different campaigns are specified in Ruedi (1995) and as approved by Sabah Parks (TS/PTD/5/4 Jld. 45 (33), TS/PTD/5/4 Jld. 47 (25)), the Economic Planning Unit (100–24/1/299), Sabah Wildlife Department (JHL: (HQ):100–42/ 1 JLD.27) and the Sabah Biodiversity Council (Ref:TK/PP:8/8Jld.2).

Molecular methods

We targeted 13 nuclear loci previously found to be single copy in the genome and informative in soricid (Esselstyn *et al.*, 2013, Igea *et al.*, 2015, Demos *et al.*, 2016) or eutherian phylogenies (Supporting Information, Table S2; Jiang *et al.*, 1998, Murphy *et al.*, 2001, Housley *et al.*, 2006). Eight markers were amplified in single reactions: six exons *ApoB* (549 bp), *BRCA1* (579 bp), *GHR10* (561 bp), *PTGER4* (471 bp), *RAG2* (444 bp) and *vWF1* (656 bp); and two introns *PRPF31* (481 bp) and *HIF1AN* (389 bp). Five other markers were amplified in multiplexed PCR reactions, including one intron (*HAT1*, 497 bp), one exon (*BDNF*, 396 bp), and three intron + exon (*GBG*, 376 bp; *DVL1*, 402 bp, and *POU2F2*, 341 bp). In addition, whole mitochondrial genomes were targeted for a subset of species that had not been previously sequenced.

DNA was extracted using phenol-chloroform and purified with an ethanol precipitation. Amicon Ultra 0.5 mL filters were used instead of ethanol precipitation to purify extracts from historic samples. Museum samples were processed in an isolated ancient DNA laboratory following strict protocols to control for contamination. Modern sample shotgun libraries were prepared as in Camacho-Sanchez *et al.* (2017), while historic ones followed the protocol described in Carøe *et al.* (2018). These libraries were quantified with a Qubit fluorometer or qPCR with Illumina Truseq dual indexes prior to equimolar pooling and sequencing.

Polymerase chain reaction (PCR) conditions are shown in Supporting Information, Table S2. In addition, we designed a set of internal primers for four of these markers (*GHR10*, 103 bp; *BDNF*, 174 bp; *BRCA*, 100 bp; and *vWF1*, 129 bp) to amplify variable areas within these loci from the highly-fragmented historic DNA. *GHR10* and *BDNF* were multiplexed, while *BRCA* and *vWF1*, which have degenerate bases and a tendency to generate dimers, were amplified in single reactions (Supporting Information, Table S2). Three replicate PCRs were run per sample in order to identify apparent mutations caused by damage or degradation of the DNA. All primers used in this study were designed with a tail to add barcodes and sequencing primers in a second round PCR (Supporting Information, Table S2).

Polymerase chain reaction (PCR) products were run on a 2% agarose gel viewed with BIO-RAD Image LabTM software (Bio-Rad Laboratories) for relative quantification of PCR products for equimolar pooling. Following this step, samples were indexed with a unique combination of forward and reverse barcodes. Dual-indexed shotgun libraries were sequenced on Illumina HiSeq 2000 or NovaSeq machines at the Genetic Resources Core Facility (GRCF) at Johns Hopkins University (Baltimore, Maryland, USA). Amplicon libraries were sequenced on a MiSeq at the same facility.

Bioinformatic methods

Preprocessing

Adaptor removal and quality trimming was performed with Trimmomatic (Bolger *et al.*, 2014) with the sliding window parameter set to 5:20, read minimum length parameter to 50 bp (Hiseq modern sample libraries) or 30 bp (Novaseq modern and historic sample libraries) and leading and trailing to 5 bp. Single quality scans were run on the raw fastq files before and after trimming with FastQC (Andrews, 2010), and viewed with MultiQC (Ewels *et al.*, 2016) enabling global trends and biases to be quickly identified.

Mitochondrial genomes

Quality trimmed reads were mapped to two reference mitochondrial genomes (*Crocidura attenuata* NC026204 or *Crocidura palawanensis* NC027243) with BWA-MEM algorithm (Li, 2013) or with Geneious Mapper iterative algorithm for up to five

iterations and medium-low sensitivity. The output of BWA was converted to BAM files with SAMtools (Li *et al.*, 2009). These were later sorted, merged and PCR duplicates removed. Finally, BAM files were imported to GENEIOUS 11.0.5, where consensus sequences were called with minimum 2x and 75% thresholds. Duplicates were removed from libraries mapped in GENEIOUS with the Dedupe plugin of the BBTools package v.35.82 (Bushnell, 2014). An alignment of our sequences with all Soricidae mitochondrial genomes available in GenBank (Supporting Information, Table S1) was made using the MAFFT v.7.450 GENEIOUS 11.0.5 plugin (Katoh & Standley, 2013) under default parameters. The control region was removed from mitogenome assemblies because it was poorly assembled in many historic samples and has been shown to provide low phylogenetic resolution and overestimation of divergence times (Duchêne *et al.*, 2011). Poorly aligned regions in rRNAs and tRNAs were trimmed with TrimAl to increase phylogenetic inference accuracy (Capella-Gutierrez *et al.*, 2009). Positions with gaps in more than 10% of the sequences (708 bp in total) were removed, leaving a final alignment of 15 360 bp. NADH dehydrogenase subunit 6 is present on the light strand and was thereby reverse complemented. Protein-coding regions were translated and inspected for frameshift mutations and for the presence of unexpected stop codons. Summary statistics of the alignment were computed with AMAS (Borowiec, 2016; Supporting Information, Table S1). Sample MNCN2979 was removed from the alignment due to suspicion of contamination during DNA extraction and the presence of spurious reads.

Genotyping of nuclear data

For nuclear sequences, we imported the trimmed reads to GENEIOUS 11.0.5 and mapped these to the closest homologous sequences found in GenBank (*Crocidura elongata*: KY771390, KY771390, KX470110, KY771782, KY772103, KY772411; *Crocidura fuliginosa*: GU981450; *Crocidura russula* (Hermann, 1780): LK936959, LK936967; *Sorex araneus* Linnaeus, 1758: NW_004545858, NW_004546023, NW_004545881; *Condylura cristata* (Linnaeus, 1758): NW_004567156) with Geneious Mapper, low-medium sensitivity and up to five iterations. We called consensus sequences in GENEIOUS with a minimum 5x coverage and 75% consensus threshold. Each locus was aligned independently with MAFFT v.7.450 GENEIOUS plugin automatic algorithm (Katoh & Standley, 2013) under default parameters. We translated exon alignments to amino acids and inspected them for insertions, deletions and premature stop codons to prevent inclusion of paralogous sequences. Amplicon sizes exceeded the read length for six loci, leaving missing data in the centre of those sequences: ApoB (40bp), BRCA1 (83bp), GHR10 (50bp), HAT1 (48bp), HF1AN (18bp) and vWF (211bp), which were trimmed. Finally, a poorly aligned region of 22 bp with a high occurrence of indels was trimmed from DVL1.

All nuclear sequences were resolved into statistically probable haplotypes using PHASE 2.1.1 (Stephens *et al.*, 2001) with an acceptance threshold of 0.70. This threshold has been shown to be optimal, given that it decreases the number of unresolved genotypes with little to no increase in false-positives (Garrick *et al.*, 2010; Demos *et al.*, 2016). The online application SeqPHASE (Flot, 2010) was used to convert FASTA files to PHASE input files, as well as to convert PHASE output back to FASTA format. Genotypes that were still unresolved following PHASE were phased manually based on the original BAM alignment (Hinckley *et al.*, 2020). Heterozygous sites were only called in historic samples when present in at least two of the three PCR replicates.

Phylogenetic analyses and divergence dating

Analyses were conducted on the following four datasets: (1) a *Cytb* dataset, including all recognized species of Sundaland and Sulawesi *Crocidura* (cytbDNA); (2) 13 nuclear loci from the main landmasses, but not all species (nDNA13); (3) a subset of six nuclear loci (ApoB, BDNF, BRCA1, GHR10, PTGER4 and vWF), including all recognized species of Sundaland and Sulawesi *Crocidura*, and most of Philippine species (nDNA6); and (4) mitochondrial genome dataset, including Malay Peninsula, Bornean and Philippine species and a reduced subset of other Sundaland/Sulawesi taxa and outgroups for divergence dating (mitoDNA).

The optimal partitioning scheme for each dataset, and substitution models for both mitochondrial maximum-likelihood analyses, were selected with ModelFinder (Kalyaanamoorthy *et al.*, 2017; Supporting Information, File S3). The number of initial partitions for mitoDNA was 53, while for cytbDNA, nDNA6 and nDNA13 protein-coding genes we input three codon positions (Supporting Information, File S3). We followed a free rate model algorithm with up to ten gamma categories for IQTREE analysis (cytbDNA and mitoDNA), and without free rate for BEAST (mitoDNA and nDNA13) analyses, BIC criteria and relaxed clustering of 10%. To save computations, only the top 10% partition schemes were considered for searching the best-fit partitioning scheme and substitution models. The best-fit partitioning scheme was the exact same seven partitions in both runs for mitoDNA, three partitions in *Cytb* and a single partition per nuclear locus (Supporting Information, File S3). For the mitoDNA dataset we performed a likelihood

ratio test (LRT) with MEGA 7 (Kumar *et al.*, 2016) to test the clock-like behaviour of the sequences, and a strict molecular clock was rejected. Given the higher evolutionary rate of the mitochondrial sequences, we tested for saturation by plotting transitions and transversions of each alignment partition against each other and against raw/uncorrected pairwise genetic distances with the dist.dna function from ape (Paradis & Schliep, 2019) in R. We also performed Xia substitution saturation tests in Dambe 7 (Xia, 2018) for each partition as a complementary line of evidence. The results were consistent between both approaches, five partitions containing third-codon positions of protein-coding genes showed high levels of saturation and were excluded, leaving two partitions of 3715 and 7864 bp for downstream divergence dating analyses (Supporting Information, File S3). We conducted phylogenetic inference through a maximum-likelihood framework with IQTREE (Nguyen *et al.*, 2015) on the cytbDNA and mitoDNA datasets. We selected an Edge-linked partition model with proportional branch lengths, and free rate model algorithm (Supporting Information, File S3). We included the Western Palaearctic *Crocidura russula* as an outgroup in these analyses. Majority rule consensus trees were generated from each analysis.

Divergence times and a species tree were estimated with mitoDNA and nDNA13 datasets in BEAST2 and *BEAST2 (Bouckaert *et al.*, 2014). The mitochondrial genome alignment was split into the best partition schemes, as identified by ModelFinder with AMAS (Borowiec, 2016). We performed site modelling through a Bayesian approach with bmodeltest (Bouckaert & Drummond, 2017), with empirical frequencies, and time-reversible models for mitoDNA or transition-transversion split for nDNA13, given that none of the ModelFinder tests for this dataset had supported GTR models. Following MEGA7 LRT tests, a lognormal (uncorrelated) relaxed clock was selected for mitoDNA, so partition site and clock models were unlinked, but only clock rates were estimated. For nDNA13, a shared strict clock with unlinked and estimated substitution rates was assumed after testing the performance of shared or unlinked relaxed clocks. Poor convergence regarding branch rate standard deviation (unlinked relaxed clock) or a branch rate standard deviation close to 0 (linked relaxed clock) suggested that rate variation among the different branches was small and that a strict clock was more appropriate (Bromham *et al.*, 2018). A birth–death process was specified in both analyses given that there is evidence for extinction due to the inclusion of extinct species in the tree (*Nesiotites hidalgo* Bate, 1945) or fossil calibrations (*C. kapsominensis* Mein & Pickford, 2006). We kept default operators and linked tree priors for the mitoDNA dataset, but kept these unlinked for the nDNA13 dataset and followed an analytical population size integration model. We removed from the initial mitoDNA alignment 8 sequences with more than 30% missing data, resulting in 83 sequences. As in Demos *et al.* (2016), for Sundaland/Sulawesi species for which samples from more than one disjunct population exist, each population was considered a terminal taxon resulting in 66 and 34 tips in mitoDNA and nDNA13, respectively. Two independent runs of Markov chains (MCMC) for Monte Carlo simulations were run for 50 000 000 (mitoDNA) and 250 000 000 generations (nDNA13), with parameters and trees sampled every 5000 generations. Convergence was checked using TRACER 1.7.1 (Rambaut *et al.*, 2018). For each run, the first 20% of sampled trees were discarded as burn-in. Two dating constraints were used to calibrate the mitoDNA phylogeny following previous Soricidae studies (Jacquet *et al.*, 2015; Bover *et al.*, 2018; Hutterer *et al.*, 2018; Chen *et al.*, 2020): (1) the split between Soricinae and Crocidurinae–Myosoricinae, estimated to have occurred about 20 Mya (Reumer, 1989; normal: mean = 20 Mya, standard deviation [SD]= 1), and (2) the oldest *Crocidura* fossil (*C. kapsominensis*), dated to 6 Mya (Mein & Pickford, 2006; lognormal: mean = 0, stdve = 1, offset = 6 Mya). Both constraints were set using hard minimum bounds and soft upper bounds, using a lognormal prior, as suggested by Parham *et al.* (2012). For the nDNA13 phylogeny, only calibration (2) could be applied. Maximum credibility consensus trees were generated from each analysis.

We estimated an additional species tree in *BEAST 2.1.1 for the nDNA6 dataset. We followed Demos *et al.* (2016) and assigned samples from separate localities as terminal taxa, resulting in 72 tips. For this six-gene analysis the loci were reduced to between three and seven sequences per lineage to keep analyses tractable and facilitate convergence (Demos *et al.*, 2016). We unlinked substitution models and trees, but linked clock models among partitions. We performed simultaneously site model averaging and phylogenetic inference with bmodeltest, selecting mutation rate estimation transition-transversion split and empirical frequencies priors. We followed an analytical population size integration model, birth–death speciation prior and assumed a strict clock for the same reasons as in nDNA13. Three independent runs were conducted for 300 000 000 generations Markov chains (MCMC) for Monte Carlo simulations, with parameters and trees sampled every 5000 generations. These runs were assessed for convergence with TRACER v.1.7.1 (Rambaut *et al.*, 2018). Maximum credibility consensus trees were generated from each analysis.

Private alleles and species delimitation

Species delimitation was performed based on mutual allelic exclusivity through haplowebs

and conspecificity matrices, as implemented in HaplowebMaker and CoMa web tools (Spöri & Flot, 2020). We considered two separate datasets of Bornean lineages: the *C. neglecta* complex, and the clade including *C. foetida*, *C. baluensis* and *C. nigripes*. Given that the haploweb approximation operates on a single locus at a time, we followed a 'conspecificity matrix' approach to combine the delimitations produced by the 13 nuclear markers and turn them into a single graphical output revealing the various hypothetical species present in the dataset. A median joining network algorithm was selected, with indels treated as a fifth state and columns with missing data masked. Haploid cytochrome *b* sequences were run separately with Haplowebmaker. GenBank sequences from the Philippine species were also included in the *C. foetida*, *C. baluensis* and *C. nigripes* analysis and *C. neglecta* from Sumatra for the second dataset. Given that ApoB did not amplify in many samples of the *C. neglecta* complex, this locus was excluded for these analyses. Sample EBD31648M was excluded from the *C. foetida* analysis because it had data for three loci; RHN38354 was excluded from PTGER and BRCA networks given the high proportion of missing data; and EBD31634M, EBD31643M, EBD31644M, MHNG1970.082, JX162658 and UMMZ174675 were excluded from the *Cytb C. neglecta* network for the same reason.

Morphological data collection

The specimens examined here (Supporting Information, Table S3) are housed in the following natural history collections: Natural History Museum, London, United Kingdom (BMNH); Estación Biológica de Doñana, Seville, Spain (EBD); Museum d'Histoire Naturelle, Geneva, Switzerland (MHNG); Museum of Comparative Zoology, Harvard University, Cambridge, Massachusetts (MCZ); Naturalis Biodiversity Center, Leiden, The Netherlands (RHN); and National Museum of Natural History, Smithsonian Institution, Washington, DC, United States (USNM). Dental eruption and wear patterns, and fusion of the basisphenoid–basioccipital sutures were checked to age each specimen. Only adults were included in the morphometric analyses. Fourteen cranial measurements were taken with electronic digital callipers to the nearest 0.01 mm by MR (Supporting Information, Table S3). Skull measurements were taken as defined in Ruedi (1995). External measurements, head–body length (HB), tail (T), hindfoot (HF), ear (E) and weight (W), were taken from specimen labels or in the field (Supporting Information, Table S3). We examined and measured the types of *C. nigripes*, *C. nigripes lipara*, *C. foetida kelabit* and *C. baluensis* (Supporting Information, Table S3). The types of *C. foetida foetida* and *C. foetida doriae* (Supporting Information, Table S1) that are hosted in Berlin could not be measured due to travel restrictions during the Coronavirus pandemic. Bivariate plots of craniodental variables were constructed based on the output of the principal component analysis (PCA). In addition, greatest skull length measurements and hair length [as in Musser *et al.* (2010)] were collected by MM, MTRH and AH from the series hosted at MCZ, USNM and EBD to test size and fur variation across an elevational gradient.

Morphometric statistical analyses

Principal component analysis. We log-transformed each measurement prior to computing the principal component analysis (PCA) so that the data was analysed on the basis of correlations instead of covariances. Principal component analysis was implemented by the prcomp command in R (R Core Team, 2018). Results were extracted and visualized with the following functions of the factoextra package (Kassambara & Mundt, 2017): fviz_pca_ind, plots PCA results; fviz_pca_biplot, biplot of individuals and variables; get_eigenvalue, extracts eigenvalues, variance percentage and cumulative variance percentage; get_pca_var, outputs each variable's contribution to variance.

Phenotypic discrimination across elevation in the clade Crocidura baluensis / C. foetida s.l. *from Sabah.* On Mt. Kinabalu, *C. foetida* has been described as distributed from lowlands up to about 1500–1600 m a.s.l., where it is replaced by *Crocidura baluensis*. Current taxonomy discriminates these species based on size and external morphological features (tail scale conspicuousness, proximal bristles in tail, relative hindfoot size and length of dorsal fur) (Thomas, 1898; Ruedi, 1995). We collected skull and external measurements (GGLS, greatest length of skull; tail; HF, hind foot) from adult specimens of *C. baluensis* and *C. foetida* from Sabah. Sex was recorded based on label information/gonad examination and age was determined based on fusion of cranial sutures. Only animals that had information on elevation were considered, leaving a total of five individuals of *C. baluensis* and 28 of *C. foetida* for the analysis (Supporting Information, Table S3). The effect of elevation was tested on each morphological trait with linear models. We included GGLS in the models to control for size on the change of HF, HB and tail, and sex to control for potential dimorphism. We also used linear models to assess the relation of hair length with elevation. Models are available at https://github.com/csmiguel/crocidura-borneo.

RESULTS

Phylogenetic analyses

Whole mitochondrial genomes were constructed from 41 individuals, generating an alignment (without control region) of 15 360 bp for 65 ingroup sequences (including some GenBank sequences), with 4487 parsimony informative sites used for IQTREE analyses. This mitogenome matrix had 10.6% missing data. A subset of 11 579 bp and 2720 parsimony informative sites with 38 ingroup taxa but only 1.5% missing data was used for the BEAST analysis (Supporting Information, Table S1). Missing data were mostly restricted to 11 sequences (MCZ36556, MCZ36564, ZRC.4.2407, ZRC.4.3437, EBD31640, EBD31643-44M, UMMZ174668, MZB16771, RHN38371 and RHN38409; Supporting Information, Table S1). Thirteen nuclear loci were sequenced from 34 individuals (Supporting Information, Table S1). Primers designed for a subset of four of the nuclear loci to amplify from historic samples worked best for those with a higher amount of DNA, which also yielded better mitogenome assemblies (Supporting Information, Table S1). Newly generated sequences were deposited in GenBank (GenBank accession numbers: MW760882–MW762425 for nuclear loci, and MW815405–MW815431 for mitochondrial genomes; Supporting Information, Table S1). All sequence alignments used in this study analyses are available at Supporting Information, File S2.

Bayesian and maximum-likelihood phylogenies yielded generally consistent topologies regardless of dataset (cytbDNA, Supporting information, Fig. S1; mitoDNA, Figs 2, 3; nDNA6, Supporting information, Fig. S2; nDNA13, Fig. 4), although there were some differences in the relationships within the *C. cf. neglecta* group and for the monophyly of Javan *C. orientalis*. The larger datasets (whole mitochondria *vs. Cytb* or 13 nuclear loci vs. 6) yielded phylogenies with higher branch support, especially for short and long branches. Divergence dates for deep nodes were similar on the mitochondrial and nuclear phylogenies, but were generally estimated as older for more recent nodes on the mitochondrial tree (Fig. 3), than on the phylogeny based on nuclear data (Fig. 4). This is expected, given that gene divergence generally occurs prior to species divergence and, therefore, mitogenome divergences are usually overestimated with regard to species tree divergences, especially for recent divergence events (Edwards & Beerli, 2000; Carstens & Knowles 2007; McCormack *et al.*, 2011). The time to the most recent common ancestor (TMRCA) of Asian *Crocidura* was estimated at 6.5 (95% HPD: 5.60–7.53; mitoDNA) and 6.1 (95% HPD: 4.11–7.25; nDNA13) million years ago (Mya), which is consistent with previous studies suggesting a Late Miocene origin (Chen *et al.*, 2020).

The species in the biogeographic region of the Philippines were monophyletic, while Wallacea (Sulawesi and the Lesser Sundas), Sundaland and Indochina *Crocidura* were not (Fig. 1). The two crossings of Wallace's Line were estimated at 6.5/6.0 Mya for the ancestor lineage of the Sulawesi radiation (mitoDNA/nDNA13) and 0.26 Mya for *C. nigripes* (nDNA13). The phylogeny shows that the ancestor of *C. nigripes* colonized Sulawesi from Borneo, not the Philippines. The radiation in Sulawesi started in the Early Pleistocene around 1.87 Mya (95% HPD: 1.33–2.44; nDNA13), after Sulawesi took on its modern form. Divergences within the Sulawesi clade were deep in comparison with the oceanic Philippine radiation, which was also highly supported as monophyletic (Giarla & Esselstyn, 2015).

Within peninsular Malaysia, *C. negligens* was nested within the diversity of *C. malayana* with high support (Figs 2, 4). Two reported specimens of *C. paradoxura* from this landmass (ZRC.4.3436 and ZRC.4.3437) were identified as *C. fuliginosa* (Supporting Information, Table S1), suggesting that *C. paradoxura* might not be present on peninsular Malaysia. The relationships of the populations currently referred to as *C. cf. neglecta* are not clear. The individual from Perlis (MHNG1970.082) was sister to the NW Borneo population (UMMZ174675) in the mitochondrial phylogenies with high support (Figs 2, 3), but sister to all other populations in nuclear phylogenies with less support (Fig. 4, PP = 0.81–0.89).

Within Sumatra, the undescribed species *Crocidura sp. nov. 2* from Mt. Singgalang (Demos *et al.*, 2016) is found to be sister to a highly divergent lineage from Mt. Tujuh (Supporting Information, File S1, Figs S1, S2), and was previously classified as *C. paradoxura* based on morphological resemblance (MZB16790, Ruedi, 1995).

Within Borneo, the species *Crocidura foetida* was paraphyletic due to the inclusion of *C. nigripes* (Fig. 4; Supporting Information, File S1, Fig. S1), *C. baluensis* (Figs 2–4; Supporting Information, File S1, Figs S1, S2) and perhaps but less likely, also the Philippine *Crocidura* clade (Figs 2, 3; Supporting Information, File S1, Figs S1, S2). North-East Borneo *C. foetida doriae* and *C. baluensis* populations were reciprocally monophyletic but showed a shallow divergence (Figs 2–4). These two taxa were sister to *C. nigripes* (Fig. 4; Supporting Information, File S1, Fig. S1) and this clade was sister to north-central/west Borneo *C. foetida foetida* and *C. foetida kelabit* with high support (Figs 2–4). Relationships within this clade, based on the nDNA6 data, were poorly supported, possibly due to a lack of variation of these markers at such a shallow scale of divergence. North-eastern Borneo lowland populations of *C. foetida s.l.* and highland parapatric *C. baluensis* diverged recently

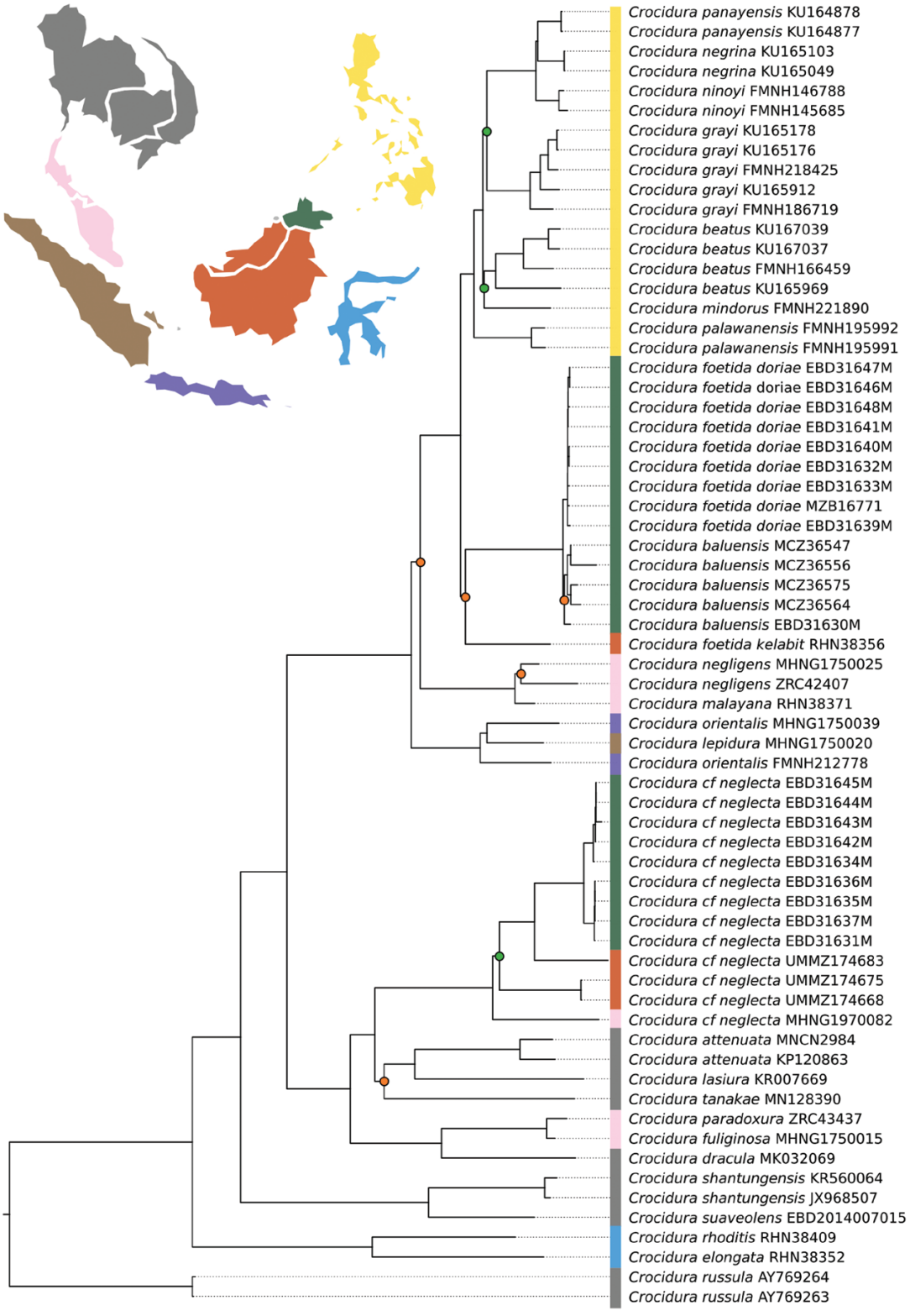

Figure 2. Mitochondrial genome maximum-likelihood consensus phylogeny of tropical east Asia *Crocidura*. Reconstructed with IQTREE. Since most nodes are highly supported [ultrafast bootstrap (UFBP) > 0.95], only less supported nodes are marked (orange circle, 0.95 < UFBP > 0.80; green circle, UFBP < 0.80). Colours on the vertical bar indicate geographical origin of sampled animals, and colours match the map. Samples from north of the Isthmus of Kra are shown in grey.

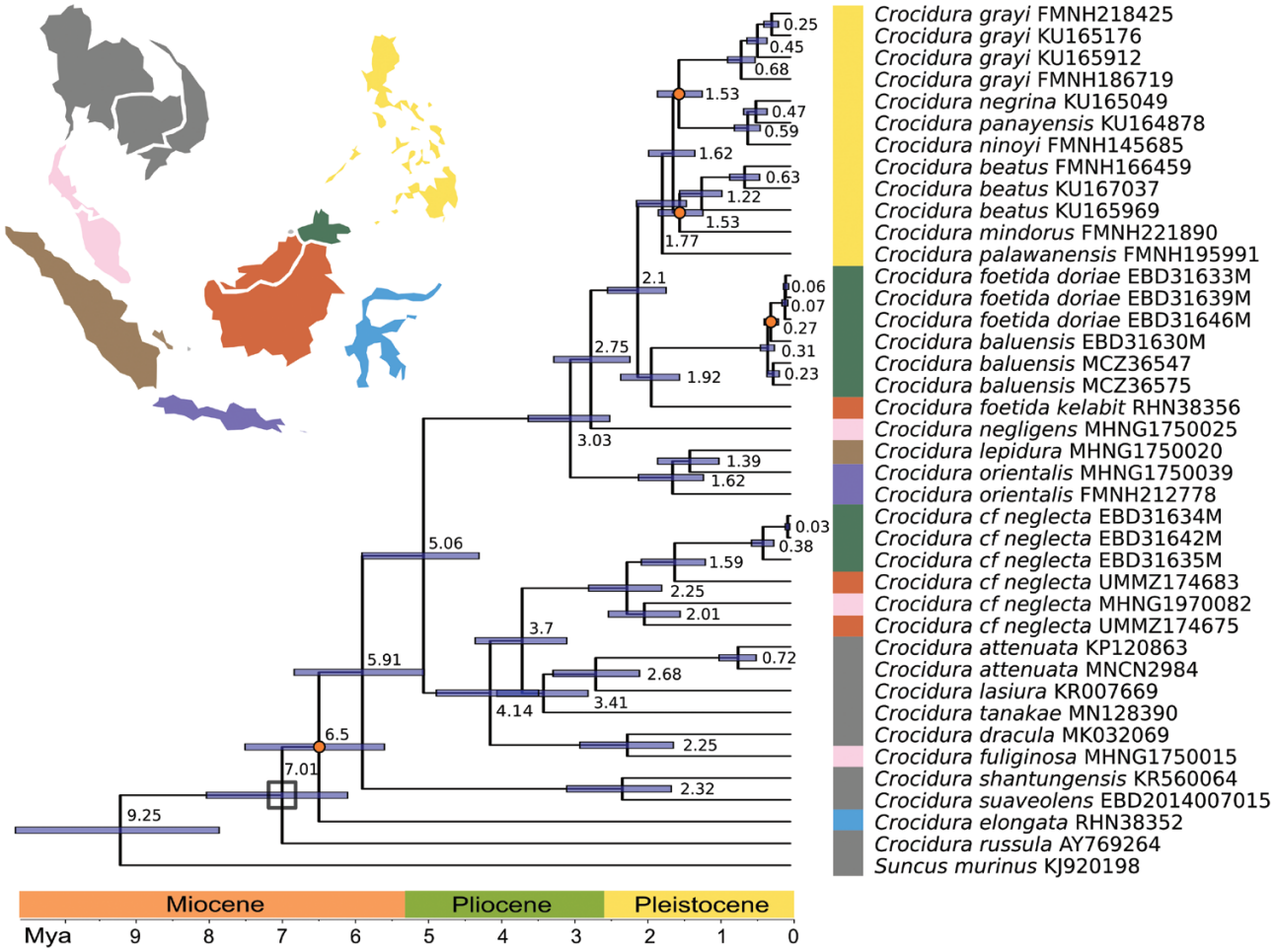

Figure 3. Bayesian maximum clade credibility tree constructed from whole mitochondrial genome sequences, with a focus on Bornean *Crocidura*. Reconstructed with BEAST2. Dates are indicated on nodes and bar on bottom. Nodes used to calibrate the dating are marked with a black square. *Uropsilus* and Soricinae outgroups were included in the phylogenetic analysis for calibration purposes but removed from the figure for clarity. Since most nodes are highly supported (PP > 0.95), only less supported nodes (0.95 < PP > 0.80) are marked with an orange circle. Colours on the vertical bar indicate geographical origin of sampled animals, and colors match the map. Samples from north of Kra are shown in grey.

from each other: 0.27 (95% HPD: 0.16–0.36, mitoDNA) and 0.11 (95% HPD: 0.01–0.19, nDNA13) Mya. The MRCA of these also diverged from that of Sulawesi *C. nigripes* populations recently, about 0.26 (95% HPD: 0.14–0.40, nDNA13) Mya.

The time of divergence among west and north-east, and central Borneo populations of *C. foetida* and *C. neglecta* clades were similar: 1.2 (95% HPD: 0.8–1.6, nuclear) and 1.9 (95% HPD: 1.5–2.3, mitochondrial) Mya for *C. foetida s.l.*; and 1.1 (95% HPD: 0.6–1.5, nuclear) and 2.2 (95% HPD: 1.8–2.8, mitochondrial) Mya for *C. neglecta s.l.* North-eastern populations of *C. cf. neglecta* had higher divergence among different mountains than *C. foetida s.l.* populations on the same mountains. Populations of *C. neglecta s.l.* from Mt. Alab (EBD31631M) and Mt. Kinabalu (EBD31635M- EBD31637M) were more closely related to each other than to those from Mt. Trusmadi (EBD31634M, EBD31642M-EBD31645M; Figs 1, 2, 4).

Private alleles and species delimitation

The *Cytb* haplotype network of the *C. neglecta* complex shows four main groups: northern Malay Peninsula, north-eastern Borneo, central Borneo and central Malay Peninsula with Sumatra and north-western Borneo (Fig. 5A). The Malay Peninsula *C. neglecta* did not share nuclear alleles with the other populations (except for DVL1, the least variable one). Central Borneo and north-western Borneo did not share alleles with the other populations at four and eight loci, while north-eastern Borneo had private alleles

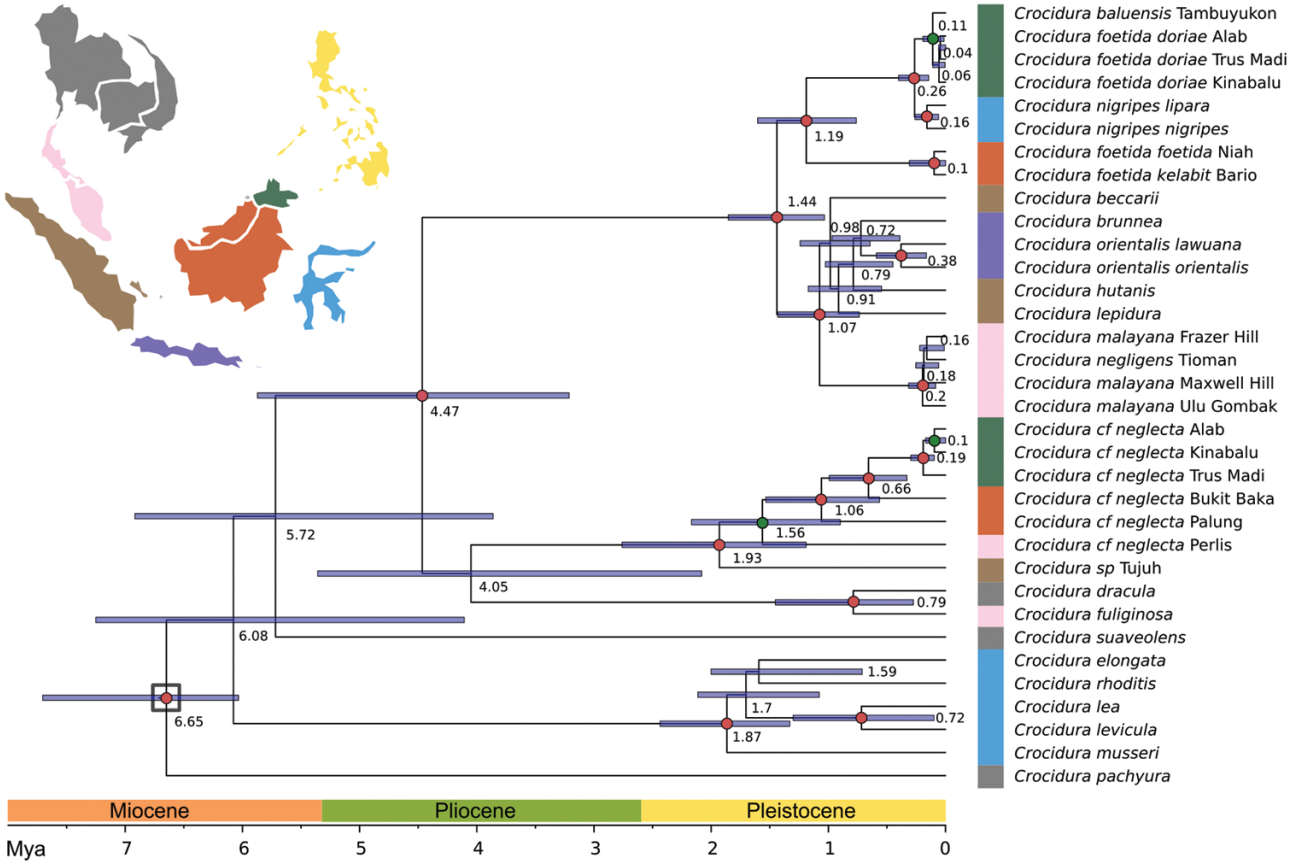

Figure 4. Species tree of Sundaland and Sulawesi *Crocidura* as estimated from 13 nuclear loci using *BEAST. Dates are indicated on nodes and bar along the bottom of phylogeny. Nodes used to calibrate the dating are marked with a black square. Highly supported nodes (PP > 0.95) are marked with a red circle, less supported nodes (0.95 < PP > 0.80) are marked with a green circle, nodes with poor support (PP < 0.80) are not marked. Samples are coloured based on its geographical origin and matching the map. Samples from north of the Isthmus of Kra are in grey.

at eight loci (Supporting Information, File S1, Fig. S3). The conspecificity matrix supported three or four separately evolving metapopulation lineages in this group. Malay Peninsula, north-eastern Borneo and north-western Borneo were highly supported, while the central Borneo population was moderately supported (Supporting Information, File S1, Fig. S4).

The *Cytb* haplotype network, which includes many more individuals from more populations than the whole mitochondrial genome phylogeny, supported the paraphyly of *C. foetida* (Fig. 5B). Populations of *C. foetida* from Sabah were closely related to *C. baluensis* and *C. nigripes* but were highly divergent from those in Sarawak (Fig. 5B). The haplotype diversity of the Philippine radiation was in between these two divergent lineages. The mid-elevation (1530 m a.s.l.) specimen from Mt. Tambuyukon was within the diversity of highland *C. baluensis*. *Crocidura foetida* from Sarawak did not share nuclear alleles with the other populations in eight of the 12 variable nuclear loci, while *C. nigripes* shared allele(s) with *C. foetida* from Sabah and *C. baluensis* in all but two markers (Supporting Information, File S1, Figs S5, S6). The highland *C. baluensis* shared allele(s) at each nuclear locus with lowland *C. foetida s.l.* from Sabah (Supporting Information, File S1, Figs S5, S6). The conspecificity matrix supported two separately evolving metapopulation lineages in this group: (1) *C. foetida s.s.* from Sarawak and (2) *C. baluensis*, *C. nigripes* and *C. foetida s.l.* from Sabah (Supporting Information, File S1, Fig. S7).

Morphometric analyses

Principal component analysis of Crocidura baluensis, C. nigripes *and* C. foetida *morphology*

A high degree of morphological differentiation among the different populations was observed (Fig. 6A) with a large part (76%) of the variance explained by the first principal component (primarily size; Supporting Information, File S3). PC1 discriminated

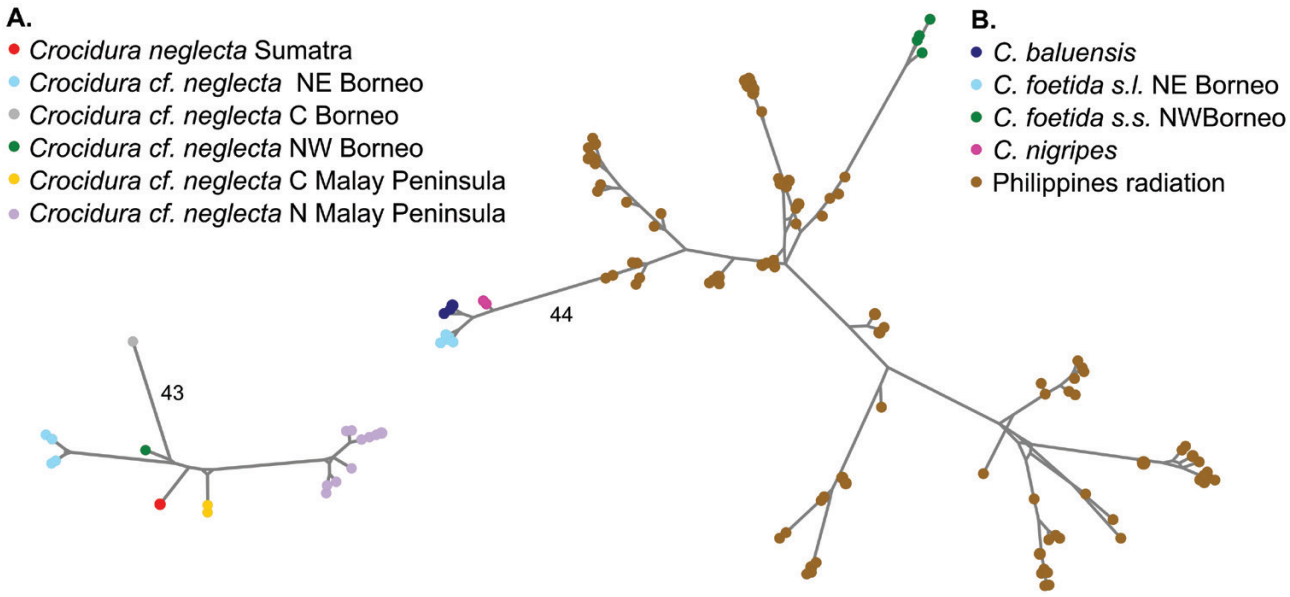

Figure 5. Whole *Cytb* haplotype median joining networks of two major lineages present in Borneo: *Crocidura neglecta* complex (A) and *C. foetida*, *C. baluensis*, *C. nigripes*, Philippine *Crocidura* (B). Branch lengths are proportional to the number of mutations, and this number has been specified on two branches as a scale.

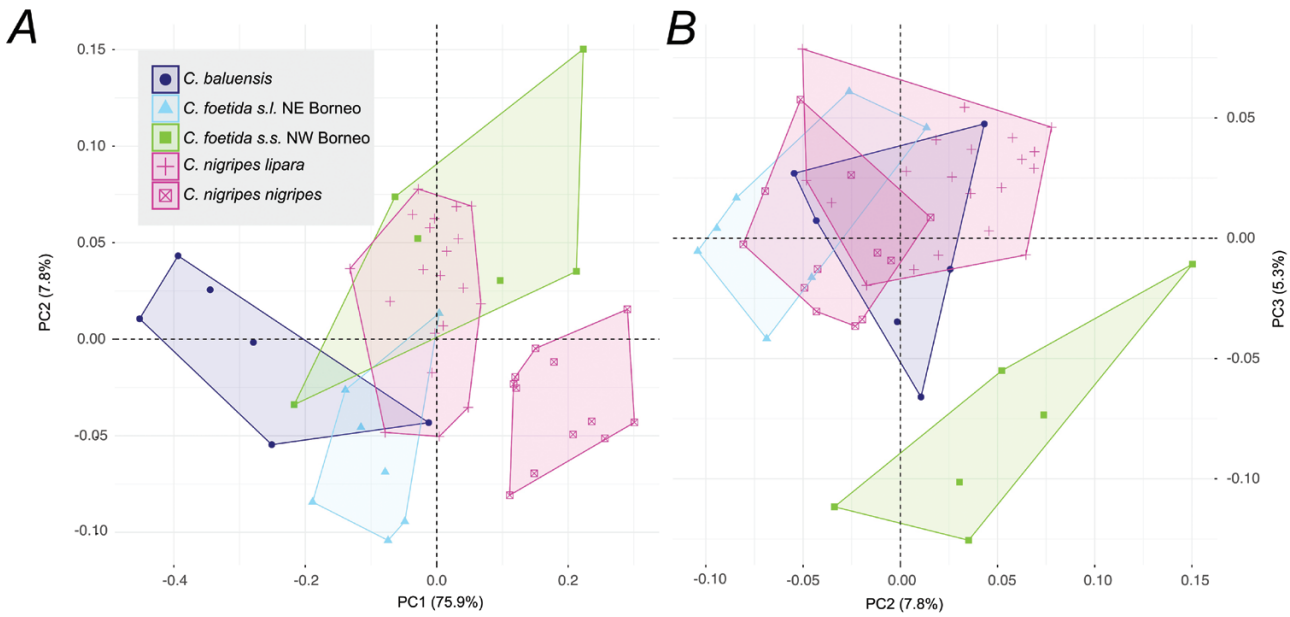

Figure 6. Morphometric variation in the *Crocidura foetida–C. baluensis-C. nigripes* complex. A, shows first and second principal components from an analysis of 14 cranial dimensions; B, shows second and third principal components of that same analysis. PC1 shows differences in size of the different populations while PCs 2 and 3 highlight differences in shape.

the largest-sized individuals (*C. baluensis*), from the medium-sized *C. foetida s.l.* and *C. nigripes lipara*, and the smallest-sized *C. nigripes nigripes*. PCs 2 and 3 explained 7.8% and 5.3% of the variance and were mainly correlated with shape. The variables with a higher contribution to PC2 and PC3 were palatal width at the third molar (53% and 20.8%) and rostral breadth (11.5% and 42%). PC2 and PC3 discriminated *C. baluensis*, *C. nigripes* and *C. foetida* s.l. north-eastern Borneo, which have a relatively gracile palate

at the rostrum level but a relatively broader palate at the third molar level, from *C. foetida* s.s. NW Borneo, which showed the opposite pattern (Fig. 6B; Supporting Information, File S3).

Phenotypic discrimination across elevation in the Crocidura baluensis / C. foetida s.l. *from Sabah clade*

Linear models showed GGLS varied positively with elevation ($R^2 = 0.36$; $P < 0.001$; Fig. 7C). After controlling for GGLS and sex, tail length and HF also increased with elevation (tail, $R^2 = 0.84$; $P < 0.001$; Fig. 7A; HF, $R^2 = 0.49$; $P = 0.001$; Fig. 7B). Linear models supported a strong positive effect of elevation on different measurements of hair length: mid-dorsum ($N = 48$; $R^2 = 0.67$; $P = 0.001$), guard hairs ($N = 28$; R2 = 0.73; $P = 0.001$), and longest guard hair ($N = 28$; $R^2 = 0.72$; $P = 0.002$) (Fig. 7D).

DISCUSSION

Permeability of the boundaries in Sundaland

Sundaland is bound by Wallace's and Huxley's lines at the edges of the Sunda Shelf between islands where there have never been land connections, and the Isthmus of Kra, a terrestrial habitat transition (Fig. 1). Few genera of small mammals are distributed across these biogeographic breaks (Groves, 2001; Esselstyn *et al.*, 2010; Rowe *et al.*, 2019), but one that is distributed across all three of these lines is the genus *Crocidura*.

Previous studies have found that the *Crocidura* of Sulawesi form a clade that radiated on that island, with one exception, *C. nigripes*, suggested to be a more recent arrival (Ruedi, 1996; Ruedi *et al.*, 1998). Mitochondrial and nuclear tree divergence dating estimated that Wallacea and Sundaland *Crocidura* split at ~6 and 7 Mya. This date is more recent than that of Sundaland-Sulawesi squirrels (~9.7–12.5 Mya;

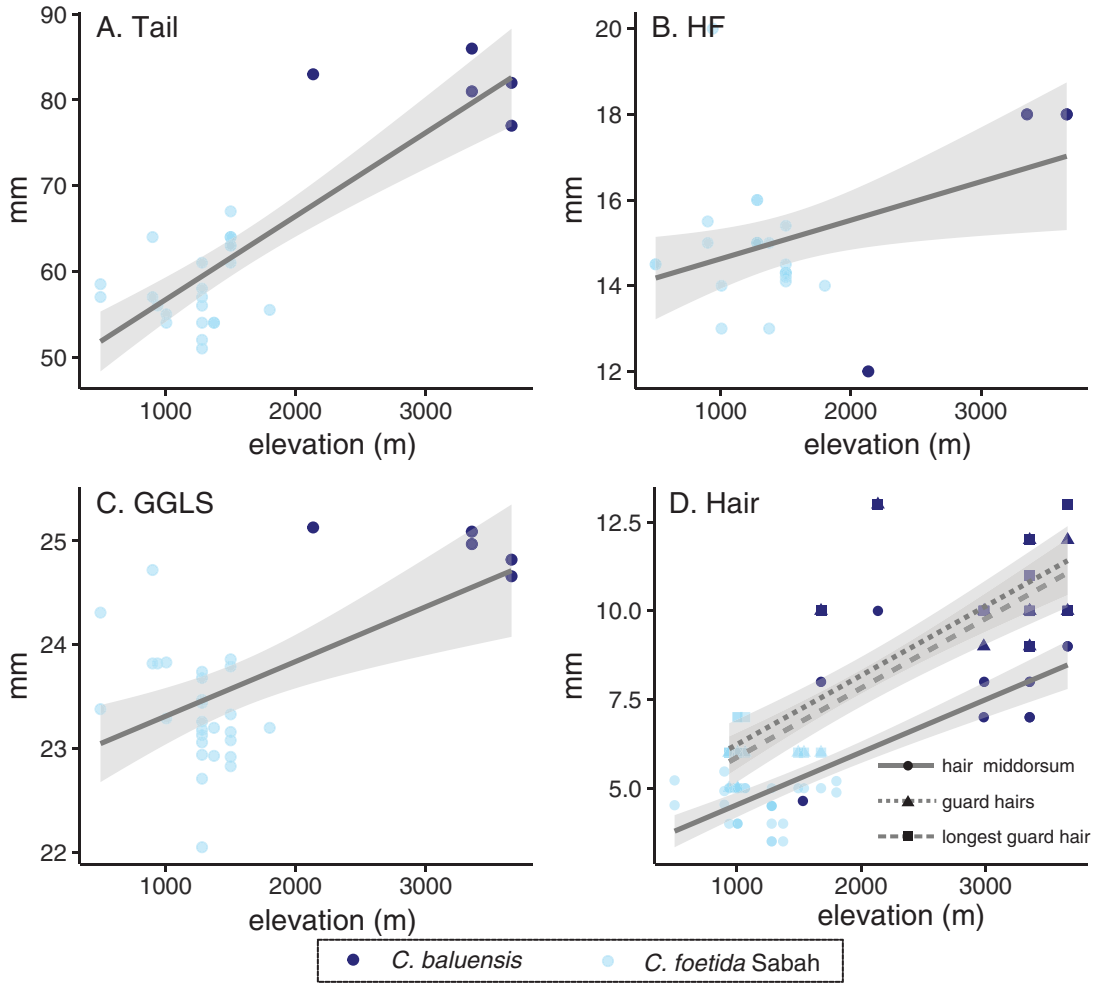

Figure 7. Relation of external morphology to elevation in the *Crocidura baluensis*/*C. foetida s.l.* clade from Sabah.

Hawkins *et al.*, 2016a) or tarsiers (~23.4 Mya; Driller *et al.*, 2015), which are more tightly associated with forest, but also older than that of Sundaland and Sulawesi spiny rats (~1.4 Mya; Rowe *et al.*, 2019), which are distributed across a wide variety of habitats. This could suggest that habitat could be an important factor in sweepstakes' dispersal events. The long gap between the divergence from Sundaland MRCA (~6–7 Mya) and diversification (~ 1.9 Mya) of Sulawesi *Crocidura* is notable and mirrors that of Sulawesi tarsiers (Driller *et al.*, 2015). Potential explanations for such a 'delayed' diversification could involve incomplete taxon sampling and/or extinction of early-branching lineages (Moen & Morlon, 2014; Driller *et al.*, 2015). The Lesser Sunda/Proto Java *C. maxi* and *C. tenuis* and Palawan *C. batakorum* were not sampled in our divergence dating analyses and might partially explain this time gap.

Our dating of the phylogeny supports a Late Pleistocene or Holocene colonization of Sulawesi from northern Borneo by the ancestor of *C. nigripes* (Fig. 4). Divergence dating overestimation is known to be pervasive in recent divergence times, so these populations possibly diverged even more recently (Ho *et al.*, 2005), perhaps within the time period that humans had begun to move around the region in boats (Van den Bergh *et al.*, 2016). Skull morphometric evidence is also consistent with a recent divergence, since *C. nigripes* exhibits the same craniodental shape as north Borneo endemic *C. baluensis* and *C. foetida s.l.* but is well differentiated from the other Bornean lineage, *C. foetida s.s*. Although a recent natural colonization cannot be discarded, as suggested for *C. tanakae* to the Batanes islands in Esselstyn & Oliveros (2010), it is also plausible that this population arrived to Sulawesi through human-mediated colonization. While there is no straightforward way to test this hypothesis with our current data, there are several lines of evidence supporting this reasoning. First, *C. foetida s.l.* is a synanthropic species, common in gardens and wastelands, making it a good candidate for human-mediated introductions (Burgin *et al.*, 2018b). Second, there is extensive evidence for multiple introduced species to Sulawesi in the literature, such as rats, porcupines, civets and deer (Groves, 2001). Third, shrews have been introduced by humans to many different islands across the world, both in recent (Bover *et al.*, 2012; Gargan *et al.*, 2016; Pinya *et al.*, 2018) and historic times (Alcover, 1980; Cheke, 1987). Further sampling in east and south Kalimantan will be needed to shed additional light on the origin of this recently diverged Sulawesi lineage.

Crocidura shrews appear to have breached Huxley's line and colonized the Philippines from Sundaland via Palawan a single time at around 1.8 Mya (Esselstyn *et al.*, 2010). Philippine and Sulawesi radiations started diversifying at the same time as Sundaland, during the Early Pleistocene (~1.6 and 1.9 Mya). This could suggest that an Early Pleistocene glacial maximum at ~1.6 Mya might have promoted dispersal events among different landmasses through a sea lowstand (200 m below present level) but also fragmented the rainforest within these landmasses, generating similar divergence patterns through two different processes: dispersal and vicariance (Abegg, 2002; Brown *et al.*, 2014). Intraspecific lineage diversification in Sundaland's red spiny rat (~1.5 Mya; Leonard *et al.*, 2015), Borneo's Low's squirrel (~1.7 Mya; Hinckley *et al.*, 2020), Sulawesi squirrels (*Prosciurillus* Ellerman, 1947; ~1.5–2.3 Mya; Hawkins *et al.*, 2016a), Philippine colugos (~1.5 Mya; Mason *et al.*, 2016) and speciation in Sunda rats (~2.2 Mya; Camacho-Sanchez *et al.*, 2017), Rusa deer (~1.8 Mya; Martins *et al.*, 2019), leaf monkeys (~1.3–1.6 Mya; *Trachypithecus* Reichenbach, 1862; Roos *et al.*, 2020), Greater Mindanao Sunda squirrels (~1.7 Mya; Den Tex *et al.*, 2010) and Philippine forest mice (*Apomys* Mearns, 1905; ~1.5 Mya; Heaney *et al.*, 2018), also took place during a similar time frame, or slightly before, such as Sulawesi tarsiers (~2.5 Mya; Driller *et al.*, 2015).

As shown for the Philippines and Wallacea (Ruedi *et al.*, 1998; Esselstyn & Oliveros, 2010), the *Crocidura* of Sundaland and Indochina were not monophyletic, indicating that the Isthmus of Kra is also a permeable boundary for this genus. In line with this high permeability, the mito-nuclear divergence dating (~ 2.25 and 0.80 Mya) between *C. fuliginosa* and *C. dracula*, might suggest secondary contact with nuclear gene flow after allopatric divergence across the Isthmus of Kra. Studies on birds have estimated similar divergence events across Kra (1.0–1.1 Mya; Lim *et al.*, 2020). Interestingly, the Sundaic-Philippine clade and Indochinese-Sundaic clade diverged during the Miocene–Pliocene boundary, at the same time as the Sundaic *Crossogale* Thomas, 1921 and Indochinese-Himalayan *Chimarrogale* Anderson, 1877 water shrews (Fig. 2; Yuan *et al.*, 2013, Abramov *et al.*, 2017; Abd Wahab *et al.*, 2020). Yuan *et al* (2013) have suggested that global warming during the Late Miocene might have isolated water shrew populations and promoted allopatric vicariance in cooler mountain climatic refugia (across mainland Asia and in northern Borneo). However, *Crocidura* might have actually expanded their distribution and colonized new areas during the rainforest expansion driven by this global warming event. Further molecular and natural history evidence will be needed to test different biogeographic scenarios.

Diversification and cryptic diversity in Sundaland

The divergence date estimates shown across Sundaland suggest that most speciation events leading to extant species took place during the Pleistocene, prior to the inundation of the Sunda Shelf 400 kya (Husson *et al.*, 2020). Other phylogeographic studies on vertebrates also support high levels of diversification when the Shelf was exposed, before 400 kya (Roberts *et al.*, 2011; Leonard *et al.*, 2015; Camacho-Sanchez, 2017; Karin *et al.*, 2017; Mason *et al.*, 2019; Cros *et al.*, 2020; Hinckley *et al.*, 2020; Lim *et al.*, 2020). Limited dispersal capabilities across non-forested ecological barriers, such as the possible 'savanna corridor', grasslands or swamps in interior Sundaland, could explain such patterns (Heaney, 1991; Cannon *et al.*, 2009; Sheldon *et al.*, 2015; Hinckley *et al.*, 2020).

Recent studies in Sumatra and Java have substantially increased the number of species in Sundaland, and decreased the distribution of individual species (Esselstyn *et al.*, 2013; Esselstyn & Achmadi, 2014; Demos *et al.*, 2016, 2017). The patterns of diversification we find, further support increased levels of endemism in Sundaland and the existence of multiple cryptic lineages that could deserve species-level recognition in the light of multiple lines of evidence (Padial *et al.*, 2010).

The complex evolutionary history of Java and lack of sampling on Mt. Lawo, the type locality of *Crocidura orientalis lawuana*, prevent us from supporting the specific status of this population. While the Rano Pani population is highly divergent from west Java populations, and more related to *C. lepidura* based on mitochondrial genome evidence, it is monophyletic and shows intraspecific levels of divergence with the nominal form based on nuclear evidence. This important divergence dating mismatch between east and west Java (1.62 and 0.38 Mya in mitoDNA and nDNA13) might be interpreted as an example of recent nuclear gene flow during montane forest expansion among separately diverging lineages (Després, 2019). Morphological and electrophoretic evidence suggest differentiation between both subspecies (Ruedi 1995, 1996), but the holotype of *C. o. lawuana* is clustered within the skull variation of the nominal form (Ruedi, 1995).

In Sumatra, our molecular evidence also supports the validity of a new cryptic species from Mt. Tujuh, previously misidentified as *C. paradoxura* due to its external resemblance (Ruedi, 1995). This population exhibits levels of differentiation from its sister-species *C. sp nov 2* from Mt. Singgalang (Demos *et al.*, 2016), which are higher than that of many recognized sister-species pairs (e.g. *C. beatus–C. ninoyi*, *C. grayi–C. mindorus*, *C. negrina–C. panayensis*, *C. hutanis–C. lepidura*, *C. beccari–C. vosmaeri* or *C. monticola–C. umbra*). However, a morphological study including the types of *C. paradoxura* and its junior synonym *C. aequicauda* (collected near our sequenced specimen and with a similar head–body/tail ratio) will have to be performed to validate and formally describe these candidate species. This pattern of endemicity on Sumatran mountains has been shown in other potential species complexes, such as *Crocidura beccari*, *Sundasciurus altitudinis* Robinson & Kloss, 1916 and *Maxomys hylomyoides* Robinson & Kloss, 1916 (Achmadi *et al.*, 2013; Demos *et al.*, 2016; Hinckley *et al.*, 2020) or recognized sister-species such as *Rattus korinchi–R. hoogerwerfi* and *C. hutanis–C. lepidura* (Ruedi, 1995; Camacho-Sanchez & Leonard, 2020). More distantly related species, such as puppet toads, parachuting frogs and montane dragons (O'Connell *et al.*, 2018; Sarker *et al.*, 2019; Shaney *et al.*, 2020), also suggest that *in situ* diversification is a generalized pattern on Sumatra and might be related to the history of these mountain forests.

Within Sulawesi, deep genetic divergences were detected across *Crocidura rhoditis*, and additional structure and polyphyly was found in *Crocidura lea s.l.* and *C. elongata s.l.* suggesting the existence of extensive cryptic diversity and an underestimation in the number of *Crocidura* species present in this island, as previously suggested (Esselstyn *et al.*, 2019).

The widespread species *C. neglecta s.l.*, exhibits high levels of diversity. The divergence among populations from Peninsular Malaysia, west Borneo and north-east Borneo is higher than that among many other recognized related species, suggesting that these might deserve specific status. Species delimitation based on private allele sharing also supports the species-level recognition of these populations and perhaps also that of the Bukit Baka population. However, we refrain from describing any of these species due to lack of morphological evidence and the small number of localities included. An increase in specimen collection and geographic coverage will be needed to characterize the fine-scale distribution of this diversity and resolve the taxonomy of this group.

Genetic structure across *C. malayana* and *C. negligens* populations on Peninsular Malaysia was subtle. Although tree topology was similar to that recovered for Malay Peninsula *C. cf. neglecta* (Omar *et al.*, 2013), with populations from Selangor sister to all the others (Perak, Tioman and Pahang), nodes were poorly supported and divergences among the different populations were similar. This scenario suggests gene flow until recent times and a simultaneous divergence, perhaps driven by Late Pleistocene habitat shifts (Camacho-Sanchez *et al.*, 2018). Isolation by distance seems an unlikely explanation for the pattern, given

that close populations such as Ulu Gombak and Frazer's Hill (~45 km) exhibit a similar divergence as Ulu Gombak and Tioman (~265 km) or Tioman and Maxwell's Hill (435 km). The shallow intraspecific-level of divergence among mainland (*C. malayana*) and Pulau Tioman (*C. negligens*) populations is in line with other species, such as rats, colugos, pangolins and mouse deer, and it suggests a recent isolation among satellite islands and nearest major landmasses populations in Sundaland (Mason *et al.*, 2019). Although the type locality of *C. negligens*, which is an offshore island in southern Thailand, has not been sampled in the molecular analyses, we synonymize *C. negligens* with *C. malayana*, since the damaged holotype of the former was included in Ruedi (1995) and morphometric classification functions showed it to be within the diversity of the Tioman and Mapor populations. This synonymization implies that a rectangular base of the palate between pterygoid processes and a bifid mesostyle of M2 are now the only characters to differentiate the sympatric and similar-sized *C. malayana* and *C. fuliginosa* (Ruedi, 1995), besides differing karyologically and genetically (Ruedi *et al.*, 1990; Ruedi & Vogel, 1995; this study).

Tail length appears to be more associated with habitat than taxonomy, with longer tails in higher elevation habitats, driving taxonomic confusion for highland populations of *C. negligens*/*C. attenuata*/*C. aequicauda* (junior synonym of *C. paradoxura*) from the Cameron Highlands (Malay Peninsula). One historically controversial specimen (ZRC43437) has a *Cytb* haplotype within the diversity of *C. fuliginosa Cytb* (Supporting Information, File S1, Fig. S1). The dark pelage coloration, tail to head body ratio and shape of the braincase of this and another specimen suggests that these are high-elevation *C. fuliginosa*, with a proportionally longer tail. Our multilocus phylogeny also supports the recent species-level recognition of *C. dracula*, previously considered a subspecies of *C. fuliginosa* (Hutterer, 2005) but recently raised to full species based on mitochondrial, chromosome and morphological evidence (Bannikova *et al.*, 2011; Burgin *et al.*, 2018a).

In situ diversification in Borneo

East–west differentiation across Borneo has been identified as a general biogeographic pattern during the last decade, with cross-taxa congruence among distantly related species, such as colugos, lesser mouse deer, pangolins, Sunda rats and Sunda squirrels (Camacho Sánchez, 2017; Hinckley *et al.*, 2020; Mason *et al.*, 2019), barbets and mountain black eyes (Den Tex & Leonard, 2013; Gawin *et al.*, 2014), frogs (Arifin *et al.*, 2011; Brown & Siler, 2014), skinks (Karin *et al.*, 2017) and trees (Ohtani *et al.*, 2013). The genetic structure shown in the two major Bornean lineages of *Crocidura* was also consistent with this pattern. West and north-east-central Borneo populations of *C. foetida* and *C. neglecta* clades not only showed the same geographic pattern, but also diverged at about the same time, during the Early Pleistocene. These shrews might have experienced the same evolutionary processes as many bird populations in the lowlands of Borneo, which were isolated during this period into rainforest refugia due to an increase in aridity and open habitats and savannas in Sundaland (Sheldon *et al.*, 2015). North-eastern Borneo populations of *C. cf. neglecta* exhibited a higher divergence among different mountains than *C. foetida s.l.*, which showed no structure, suggesting recent or ongoing gene flow. *Crocidura cf. neglecta* populations from Mt. Alab and Mt. Kinabalu were more closely related to each other than to Mt. Trusmadi. Although Mt. Alab is equidistant between Mt. Kinabalu and Mt. Trusmadi, (~35 km), and the entire area surrounding these populations is mountainous, such differentiation could be explained by the presence of a higher montane habitat corridor connecting the former localities and the existence of a possibly less suitable, low-montane habitat matrix isolating the latter (~900 m) (Camacho-Sanchez *et al.*, 2018). While *C. cf. neglecta* has been recorded in the lowlands of north-western Borneo, it has only been collected over 1500 m a.s.l. in north-eastern Borneo. Similarly, within *C. foetida s.s.*, the population from Niah was more closely related to the one in Bario than to Bintulu (USNM590298-9; Supporting Information, File S1, Fig. S2). This is interesting because Niah is almost twice as close to Bintulu and at a similar low elevation, while Bario is on the other side of one of the largest rivers in Borneo, the Baram, and up in the Kelabit Highlands.

The poorly known Bornean endemic *C. baluensis* seems to represent a high-elevation adapted lineage recently diverged from lowland *C. foetida s.l.* from Sabah, and not a long-distance immigrant and pre-adapted highland lineage related to *C. lepidura* from Sumatra, as previously thought due to its similar large size and pelage (Ruedi, 1995). Therefore, the morphological resemblance to the latter might be due to convergent evolution driven by strong selective pressures imposed by elevation. Similar remarkable morphological convergence associated with high elevation has been shown among the Bornean mountain rat *R. baluensis* and Sumatran *R. korinchi* and *R. hoogerwerfi* Chasen, 1939 (Musser, 1986; Camacho-Sanchez & Leonard, 2020) or among Asian mainland plain long-nosed squirrels (*Dremomys* Heude, 1898) and Bornean mountain ground squirrel (*Sundasciurus everetti*) (Hawkins *et al.*, 2016b). Merckx *et al.* (2015) showed that most Mt. Kinabalu endemics had a recent origin, and defined as centric

endemics those that have local sister-taxa in Bornean lowlands, occur on average at lower elevations (mean lower elevational boundary, 1724 m; s.d., 728 m), tend to have lower dispersal capacities and are frequently not strict Mt. Kinabalu endemics, as they are also found on other mountains around Mt. Kinabalu. This seems to be the case for *C. baluensis*, which recently evolved from a local, lowland ancestor and we show here is also present on nearby Mt. Tambuyukon.

The morphological study of a series of shrews from northern Borneo suggests that body size, relative hindfoot length, tail length and hair length increase with elevation in *C. foetida s.l./C. baluensis*. An increase in hair length with elevation has also been shown in mountain treeshrews of Borneo (*Tupaia montana*) and New Guinea rats (*Rattus spp.*) (Taylor *et al.*, 1985; Hinckley, 2021). Mountain treeshrew size decreased from the lowest elevations (< 1000 m) to middle elevations (2000–2500 m) and then increased again from middle elevations to highest elevations (Hinckley, 2021). The body size variation of these white-toothed shrews might have a similar pattern, but improved sampling of *C. foetida s.l.* at lower elevations (< 1000 m) and between 2000 and 3000 m is necessary to validate this pattern. The relative hindfoot increase shown along elevation (Fig. 7B) is consistent with the pattern found in mountain treeshrews, and also Bornean bird communities (Boyce *et al.*, 2019), while the tail increase is not (Fig. 7A; Hinckley, 2021). Hindfoot size has been hypothesized to be related with arboreality in the tropics (Alroy, 2019), while tail length has been shown to be related with scansoriality in shrews (Hutterer, 1985), squirrels (Hayssen, 2008), cricetids (Kingsley *et al.*, 2017), murines (Nations *et al.*, 2019) and treeshrews (Martin, 1968). Thus, differences in the relative length of these might imply different locomotion and foraging strategies associated with elevation or in the different habitats. Two other high-elevation populations in this study might mirror this pattern: *C. fuliginosa* from the Cameron Highlands and *C. foetida kelabit* from the Kelabit Highlands. Our molecular and skull morphometric evidence position *C. foetida kelabit* within the diversity of *C. foetida s.s*. The other subspecies, *Crocidura foetida doriae*, is distinguished from the nominal form based on two traits: size and pelage (Ruedi, 1995). The validity of this subspecies was questioned by Medway (1977) who synonymized it after attributing to extreme attrition the dental peculiarities that distinguish the nominal form from it. Our results suggest that pelage length and size are also closely linked to elevation and labile traits. The lack of a specific type locality for this subspecies, the limited genetic data we were able to collect from the holotype (a single uninformative fragment) and the fact that we could not examine it, complicates taxonomic revisions of *C. f. doriae*.

The molecular evidence found that *Crocidura foetida* is paraphyletic due to the inclusion of *C. nigripes* and *C. baluensis*. Species delimitation based on private alleles supported that north-east Borneo populations of *C. foetida s.l.*, together with *C. nigripes* and *C. baluensis*, represent a distinct, separately evolving lineage to the ones from north-west–north-central Borneo: *C. foetida s.s.* True *C. foetida* shows a distinct skull shape as compared to NE *C. foetida s.l.*, *C. baluensis* and *C. nigripes*, which largely overlap in shape. Thomas (1898) gave in the original description two qualitative characters that distinguish *C. baluensis* from *C. foetida*: the almost complete lack of bristle hairs at the base of the tail and the long mid-dorsal fur (8–10 mm in *C. baluensis*, much longer than *C. foetida*). Ruedi (1995) also suggested conspicuous, coarse tail-scales as an additional feature of *C. baluensis*. The expanded specimen series examined in this study suggests these qualitative characters may be related to elevational ecophenotypic variation, as found in sympatric mountain treeshrews (Parker *et al.*, 2020; Hinckley, 2021), although key sampling at middle elevations is missing. *Crocidura baluensis* specimens seem to generally have a low number of proximal tail bristles (~2–3, although subadults have ~5–10), coarse scales in some individuals, a darker fore and hindfoot coloration and conspicuous locks of blond bristles over hindfoot nails. The lowland lineage '*C. foetida s.l.*' has a higher number of proximal tail bristles (4 to more than ~10), no coarse scales, lighter limbs and inconspicuous or lack of blond bristles over hindfoot nails (Supporting Information, Fig. S8). However, high-elevation '*C. foetida s.l.* lineage' specimens from Mt. Alab (1800–1950 m) seem to show '*C. baluensis*' features [three and six proximal bristles, overlapping with the range of *C. baluensis*, coarse tail scales, dark brown limbs and prominent (adult) or less prominent but conspicuous (subadult) lock of blond bristles]. The *C. baluensis* specimen from Mt. Tambuyukon showed a much shorter hair length (~4.5 mm) than higher elevation specimens from the same mitochondrial lineage (7–10 mm), more concordant with the length of hair from lowland specimens (3.5–5.5 mm). Unfortunately, the *C. baluensis* specimen is a subadult, so it could not be included in the size comparison. Unless this specimen represents an admixed individual, this discrete character (hair) that defined *C. baluensis* could also be related to elevation.

CONCLUSIONS

We have constructed a multilocus phylogeny of *Crocidura* from all major Sundaland landmasses, the Philippines and Sulawesi to address the biogeography of this group. Our phylogenies support:

(1) most speciation events leading to the extant species took place during the Pleistocene, prior to the inundation of the Sunda Shelf at ~400 kya; (2) Borneo as a centre of *in situ* diversification; (3) the pattern of divergence between populations in east and west Borneo found in other taxa is also present in both widespread lineages of Bornean *Crocidura*; and (4) the genus *Crocidura* contains several undescribed species in Sundaland and Sulawesi, and their description will change the pattern of species distributions from few widespread to more restricted distributions.

ACKNOWLEDGEMENTS

We thank R. Intang and C. Baidis who assisted us during fieldwork. Logistic support was provided by L. Gonsilou and D. Hinckley and A. Gorog provided samples. Ana Piriz and Irene Quintanilla provided important advice during laboratory work and support during amplicon and library pooling. Sabah Parks, Sabah Forestry Department and Sabah Biodiversity Council provided research permits and logistic support. Access to tissue samples was granted by N. Duncan, S. Katanova, M. Omura, B. Zimkus, K. Lim, M. Chua, F. Mayer, S. Bock, I. Rey, A. Garvia and O. Glaizot hosted at AMNH, MCZ Harvard, LKC, Berlin MFN, MNCN and MZ Lausanne. Important pictures and/or measurements of sequenced specimens were provided by K. Lim and M. Omura who hosted these in their collections. Assistance at EBD-CSIC scientific collection was provided by C. Urdiales. David Aragonés from the GIS and Remote Sensing Laboratory (LAST-EBD) of the Doñana Biological Station, CSIC made the map. Logistical support was provided by Laboratorio de Ecología Molecular and Laboratorio de Ecología Acuática, Estación Biológica de Doñana, CSIC (LEM-EBD) and Doñana ICTS-RBD.

AH was supported by a Ministerio de Economía y Competitividad contract CGL2014-58793-P. The Ministerio de Ciencia e Innovación grant CGL2017-86068-P to JAL funded this study. We acknowledge support of the publication fee by the CSIC Open Access Publication Support Initiative through its Unit of Information Resources for Research (URICI).

REFERENCES

Abd Wahab MF, **Pathmanathan D**, **Motokawa M**, **Khan FAA**, **Omar H. 2020.** Taxonomic assessment of the Malayan water shrew *Chimarrogale hantu* Harrison, 1958 and reclassification to the genus *Crossogale*. *Mammalian Biology* **100:** 399–409.

Abegg C. 2002. Macaque evolution and dispersal in insular south-east Asia. *Biological Journal of the Linnean Society* **75:** 555–576.

Abramov AV, **Jenkins PD**, **Rozhnov VV**, **Kalinin AA. 2008.** Description of a new species of *Crocidura* (Soricomorpha: Soricidae) from the island of Phu Quoc, Vietnam. *Mammalia* **72:** 269–272.

Abramov AV, **Bannikova AA**, **Lebedev VS**, **Rozhnov VV. 2017.** Revision of *Chimarrogale* (Lipotyphla: Soricidae) from Vietnam with comments on taxonomy and biogeography of Asiatic water shrews. *Zootaxa* **4232:** 216–230.

Achmadi AS, **Esselstyn JA**, **Rowe KC**, **Maryanto I**, **Abdullah MT. 2013.** Phylogeny, diversity, and biogeography of Southeast Asian spiny rats (*Maxomys*). *Journal of Mammalogy* **94:** 1412–1423.

Alcover JA. 1980. Note on the origin of the present mammalian fauna from the Balearic and Pityusic islands. *Miscellània Zoològica* **6:** 141–149.

Alroy J. 2019. Small mammals have big tails in the tropics. *Global Ecology and Biogeography* **28:** 1042–1050.

Andrews S. 2010. *FastQC: a quality control tool for high throughput sequence data*. Available at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc. Accessed April 2020.

Arifin U, **Iskandar DT**, **Bickford DP**, **Brown RM**, **Meier R**, **Kutty SN. 2011.** Phylogenetic relationships within the genus *Staurois* (Anura, Ranidae) based on 16S rRNA sequences. *Zootaxa* **2744:** 39–52.

Bannikova AA, **Abramov AV**, **Borisenko AV**, **Lebedev VS**, **Rozhnov VV. 2011.** Mitochondrial diversity of the white-toothed shrews (Mammalia, Eulipotyphla, *Crocidura*) in Vietnam. *Zootaxa* **2812:** 1–20.

Bolger AM, **Lohse M**, **Usadel B. 2014.** Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30:** 2114–2120.

Borowiec ML. 2016. AMAS: a fast tool for alignment manipulation and computing of summary statistics. *PeerJ* **4:** e1660.

Bouckaert RR, **Drummond AJ. 2017.** bModelTest: Bayesian phylogenetic site model averaging and model comparison. *BMC Evolutionary Biology* **17:** 42.

Bouckaert RR, **Heled J**, **Kühnert D**, **Vaughan T**, **Wu CH**, **Xie D**, **Suchard MA**, **Rambaut A**, **Drummond AJ. 2014.** BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Computational Biology* **10:** e1003537.

Bover P, **Parpal L**, **Pons J**, **Alcover JA. 2012.** Evidence for a recent introduction of *Crocidura russula* (Mammalia, Eulipotyphla, Soricomorpha) in Mallorca (Balearic Islands, western Mediterranean Sea). *Mammalia* **76:** 113–116.

Bover P, **Mitchell KJ**, **Llamas B**, **Rofes J**, **Thomson VA**, **Cuenca-Bescós G**, **Alcover JA**, **Cooper A**, **Pons J. 2018.** Molecular phylogenetics supports the origin of an endemic Balearic shrew lineage (*Nesiotites*) coincident with the Messinian Salinity Crisis. *Molecular Phylogenetics and Evolution* **125:** 188–195.

Boyce AJ, **Shakya S**, **Sheldon FH**, **Moyle RG**, **Martin TE. 2019.** Biotic interactions are the dominant drivers of

phylogenetic and functional structure in bird communities along a tropical elevational gradient. *The Auk* **20:** 1–14.

Bromham L, **Duchêne S**, **Hua X**, **Ritchie AM**, **Duchêne DA**, **Ho SY. 2018.** Bayesian molecular dating: opening up the black box. *Biological Reviews* **93:** 1165–1191.

Brown RM, **Siler CD. 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 RM, **Weghorst JA**, **Olson KV**, **Duya MR**, **Barley AJ**, **Duya MV**, **Shekelle M**, **Neri-Arboleda I**, **Esselstyn JA**, **Dominy NJ**, **Ong PS. 2014.** Conservation genetics of the Philippine tarsier: cryptic genetic variation restructures conservation priorities for an island archipelago primate. *PLoS One* **9:** e104340.

Burgin C, **He K**, **Haslauer R**, **Sheftel B**, **Jenkins P**, **Ruedi M**, **Hintsche S**, **Motokawa M**, **Hinckley A**, **Hutterer R. 2018a.** Family Soricidae (Shrews). In: Wilson, DE, Mittermeier RA, eds. *Handbook of the mammals of the world – Vol. 8: insectivores, sloths and colugos.* Barcelona: Lynx Edicions, 332–551.

Burgin CJ, **Colella JP**, **Kahn PL**, **Upham NS. 2018b.** How many species of mammals are there? *Journal of Mammalogy* **99:** 1–14.

Bushnell B. 2014. *BBTools software package.* Available at: http://sourceforge. net/projects/bbmap. Accessed April 2020.

Camacho-Sánchez M. 2017. *Evolution in Sundaland: insights from comparative phylogeography of* Rattus *and* Sundamys *rats.* Unpublished D. Phil. Thesis, University Pablo de Olavide.

Camacho-Sanchez M, **Leonard JA. 2020.** Mitogenomes reveal multiple colonization of mountains by *Rattus* in Sundaland. *Journal of Heredity* **111:** 392–404.

Camacho-Sanchez M, **Leonard JA**, **Fitriana Y**, **Tilak MK**, **Fabre PH. 2017.** The generic status of *Rattus annandalei* (Bonhote, 1903)(Rodentia, Murinae) and its evolutionary implications. *Journal of Mammalogy* **98:** 1340–1355.

Camacho-Sanchez M, **Quintanilla I**, **Hawkins MT**, **Tuh FY**, **Wells K**, **Maldonado JE**, **Leonard JA. 2018.** Interglacial refugia on tropical mountains: novel insights from the summit rat (*Rattus baluensis*), a Borneo mountain endemic. *Diversity and Distributions* **24:** 1252–1266.

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 USA* **106**: 11188–11193.

Capella-Gutiérrez S, **Silla-Martínez JM**, **Gabaldón T. 2009.** TrimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25:** 1972–1973.

Carøe C, **Gopalakrishnan S**, **Vinner L**, **Mak SS**, **Sinding MHS**, **Samaniego JA**, **Wales N**, **Sicheritz-Pontén T**, **Gilbert MTP. 2018.** Single-tube library preparation for degraded DNA. *Methods in Ecology and Evolution* **9:** 410–419.

Carstens BC, **Knowles LL. 2007.** Estimating species phylogeny from gene-tree probabilities despite incomplete lineage sorting: an example from *Melanoplus* grasshoppers. *Systematic Biology* **56:** 400–411.

Cheke AS. 1987. An ecological history of the Mascarene Islands, with particular reference to extinctions and introductions of land vertebrates. In: Diamond AS (ed.). *Studies of Mascarene birds*. Cambridge: Cambridge University Press, 5–89.

Chen S, **Qing J**, **Liu Z**, **Liu Y**, **Tang M**, **Murphy RW**, **Pu Y**, **Wang X**, **Tang K**, **Guo K**, **Jiang X. 2020.** Multilocus phylogeny and cryptic diversity of white-toothed shrews (Mammalia, Eulipotyphla, *Crocidura*) in China. *BMC Evolutionary Biology* **20:** 1–14.

Cros E, **Chattopadhyay B**, **Garg KM**, **Ng NS**, **Tomassi S**, **Benedick S**, **Edwards DP**, **Rheindt FE. 2020.** Quaternary land bridges have not been universal conduits of gene flow. *Molecular Ecology* **29:** 2692–2706.

Davison GWH. 1984. New records of P. Malaysian and Thai shrews. *Malayan Nature Journal* **36:** 211–215.

De Bruyn M, **Stelbrink B**, **Morley RJ**, **Hall R**, **Carvalho GR**, **Cannon CH**, **van den Bergh G**, **Meijaard E**, **Metcalfe I**, **Boitani L**, **Maiorano L. 2014.** Borneo and Indochina are major evolutionary hotspots for Southeast Asian biodiversity. *Systematic Biology* **63:** 879–901.

Demos TC, **Achmadi AS**, **Giarla TC**, **Handika H**, **Rowe KC**, **Esselstyn JA 2016.** Local endemism and within-island diversification of shrews illustrate the importance of speciation in building Sundaland mammal diversity. *Molecular Ecology* **25:** 5158–5173.

Demos TC, **Achmadi AS**, **Handika H**, **Rowe KC**, **Esselstyn JA. 2017.** A new species of shrew (Soricomorpha: *Crocidura*) from Java, Indonesia: possible character displacement despite interspecific gene flow. *Journal of Mammalogy* **98:** 183–193.

Den Tex RJ, **Leonard JA. 2013.** A molecular phylogeny of Asian barbets: speciation and extinction in the tropics. *Molecular Phylogenetics and Evolution* **68:** 1–13.

Den Tex RJ, **Thorington RW**, **Maldonado JE**, **Leonard JA. 2010.** Speciation dynamics in the SE Asian tropics: putting a time perspective on the phylogeny and biogeography of Sundaland tree squirrels, *Sundasciurus. Molecular Phylogenetics and Evolution* **55:** 711–720.

Després L. 2019. One, two or more species? Mitonuclear discordance and species delimitation. *Molecular Ecology* **28:** 3845–3847.

Driller C, **Merker S**, **Perwitasari-Farajallah D**, **Sinaga W**, **Anggraeni N**, **Zischler H. 2015.** Stop and go—waves of tarsier dispersal mirror the genesis of Sulawesi island. *PLoS One* **10:** e0141212.

Duchêne, S, **Archer FI**, **Vilstrup J**, **Caballero S**, **Morin PA. 2011.** Mitogenome phylogenetics: the impact of using single regions and partitioning schemes on topology, substitution rate and divergence time estimation. *PLoS One* **6:** e27138.

Edwards SV, **Beerli P. 2000.** Perspective: gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. *Evolution* **54:** 1839–1854

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

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

Esselstyn JA, **Oliveros CH. 2010.** Colonization of the Philippines from Taiwan: a multi-locus test of the biogeographic and phylogenetic relationships of isolated populations of shrews. *Journal of Biogeography* **37:** 1504–1514.

Esselstyn JA, **Oliveros CH**, **Moyle RG**, **Peterson AT**, **McGuire JA**, **Brown RM. 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**, **Siler CD**, **Evans BJ. 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**, **Handika H**, **Giarla TC**, **Rowe KC. 2019.** A new climbing shrew from Sulawesi highlights the tangled taxonomy of an endemic radiation. *Journal of Mammalogy* **100:** 1713–1725.

Ewels P, **Magnusson M**, **Lundin S**, **Käller M. 2016.** MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32:** 3047–3048.

Flot JF. 2010. SeqPHASE: a web tool for interconverting PHASE input/output files and FASTA sequence alignments. *Molecular Ecology Resources* **10:** 162–166.

Gargan LM, **Cornette R**, **Yearsley JM**, **Montgomery WI**, **Paupério J**, **Alves PC**, **Butler F**, **Pascal M**, **Tresset A**, **Herrel A**, **Lusby J. 2016.** Molecular and morphological insights into the origin of the invasive greater white-toothed shrew (*Crocidura russula*) in Ireland. *Biological Invasions* **18:** 857–871.

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.

Gawin DF, **Rahman MA**, **Ramji MFS**, **Smith BT**, **Lim HC**, **Moyle RG**, **Sheldon FH. 2014.** Patterns of avian diversification in Borneo: the case of the endemic mountain black-eye (*Chlorocharis emiliae*). *The Auk: Ornithological Advances* **131:** 86–99.

Giarla TC, **Esselstyn JA. 2015.** The challenges of resolving a rapid, recent radiation: empirical and simulated phylogenomics of Philippine shrews. *Systematic Biology* **64:** 727–740.

Groves, C. 2001. Mammals in Sulawesi: where did they come from and when, and what happened to them when they got there. In: Metcalfe I, Smith JM, Morwood M, Davidson I. *Faunal and floral migrations and evolution in SE Asia-Australia.* Boca Raton: CRC press, 333–342.

Hall R. 2013. The palaeogeography of Sundaland and Wallacea since the Late Jurassic. *Journal of Limnology* **72:** 1–17.

Hawkins MT, **Leonard JA**, **Helgen KM**, **McDonough MM**, **Rockwood LL**, **Maldonado JE. 2016a.** Evolutionary history of endemic Sulawesi squirrels constructed from UCEs and mitogenomes sequenced from museum specimens. *BMC Evolutionary Biology* **16:** 1–16.

Hawkins MT, **Helgen KM**, **Maldonado JE**, **Rockwood LL**, **Tsuchiya MT**, **Leonard JA. 2016b.** Phylogeny, biogeography and systematic revision of plain long-nosed squirrels (genus *Dremomys*, Nannosciurinae). *Molecular Phylogenetics and Evolution* **94:** 752–764.

Hayssen V. 2008. Patterns of body and tail length and body mass in Sciuridae. *Journal of Mammalogy* **89:** 852–873.

Heaney LR. 1984. Mammalian species richness on islands on the Sunda Shelf, Southeast Asia. *Oecologia* **61:** 11–17.

Heaney LR. 1991. A synopsis of climatic and vegetational change in Southeast Asia. *In Tropical Forests and Climate* **19:** 53–61.

Heaney LR, **Kyriazis CC**, **Balete DS**, **Steppan SJ**, **Rickart EA. 2018.** How small an island? Speciation by endemic mammals (*Apomys*, Muridae) on an oceanic Philippine island. *Journal of Biogeography* **45:** 1675–1687.

Hinckley A. 2021. *Habitat driven diversification in small mammals of tropical east Asia*. University of Seville.

Hinckley A, **Hawkins M**, **Achmadi A**, **Maldonado JE**, **Leonard JA. 2020.** Ancient divergence driven by geographic isolation and ecological adaptation in forest dependent Sundaland tree squirrels. *Frontiers in Ecology and Evolution* **8:** 208.

Hinckley A, **Sanchez-Donoso I**, **Comas M**, **Camacho-Sanchez M**, **Hasan NH**, **Leonard JA. 2021.** Challenging ecogeographical rules: phenotypic variation in the mountain treeshrew (*Tupaia montana*) along tropical elevational gradients. In review.

Ho SY, **Phillips MJ**, **Cooper A**, **Drummond AJ. 2005.** Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. *Molecular Biology and Evolution* **22:** 1561–1568.

Housley DJ, **Zalewski ZA**, **Beckett SE**, **Venta PJ. 2006.** Design factors that influence PCR amplification success of cross-species primers among 1147 mammalian primer pairs. *BMC Genomics* **7:** 253.

Husson L, **Boucher FC**, **Sarr AC**, **Sepulchre P**, **Cahyarini SY. 2020.** Evidence of Sundaland's subsidence requires revisiting its biogeography. *Journal of Biogeography* **47:** 843–853.

Hutterer R. 1985. Anatomical adaptations of shrews. *Mammal Review* **15:** 43–55.

Hutterer R. 2005. Order Soricomorpha. In: Wilson DE, Reeder DM, eds. *Mammal species of the world: a taxonomic and geographic reference*, *Vol.* **1**. Baltimore: John Hopkins University Press, 220–311.

Hutterer R, **Balete DS**, **Giarla TC**, **Heaney LR**, **Esselstyn JA. 2018.** A new genus and species of shrew (Mammalia: Soricidae) from Palawan Island, Philippines. *Journal of Mammalogy* **99:** 518–536.

Igea J, **Aymerich P**, **Bannikova AA**, **Gosálbez J**, **Castresana J. 2015.** Multilocus species trees and species delimitation in a temporal context: application to the water shrews of the genus *Neomys*. *BMC Evolutionary Biology* **15:** 209.

Jacquet F, **Denys C**, **Verheyen E**, **Bryja J**, **Hutterer R**, **Peterhans JCK**, **Stanley WT**, **Goodman SM**,

Couloux A, **Colyn M**, **Nicolas V. 2015.** Phylogeography and evolutionary history of the *Crocidura olivieri* complex (Mammalia,Soricomorpha): from a forest origin to broad ecological expansion across Africa. *BMC Evolutionary Biology* **15:** 71.

Jenkins PD. 1976. Variation in Eurasian shrews of the genus *Crocidura* (Insectivora: Soricidae). *Bulletin of the British Museum of Natural History (Zoology)* **30:** 271–309.

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

Jenkins PD, **Abramov AV**, **Rozhnov VV**, **Makarova OV. 2007.** Description of two new species of white-toothed shrews belonging to the genus *Crocidura* (Soricomorpha: Soricidae) from the Ngoc Linh Mountain, Vietnam. *Zootaxa* **1589:** 57–68.

Jenkins PD, **Lunde DP**, **Moncrieff CB. 2009.** Chapter 10. Descriptions of new species of *Crocidura* (Soricomorpha: Soricidae) from mainland Southeast Asia, with synopses of previously described species and remarks on biogeography. In: Voss RS, Carleton MC eds. Systematic mammalogy: contributions in honour of Guy G. Musser. *Bulletin of the American Museum of Natural History* **331:** 356–405.

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.

Jiang Z, **Priat C**, **Galibert F. 1998.** Traced orthologous amplified sequence tags (TOASTs) and mammalian comparative maps. *Mammalian Genome* **9:** 577–587.

Kalyaanamoorthy S, **Minh BQ**, **Wong TK**, **von Haeseler A**, **Jermiin LS. 2017.** ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods* **14:** 587.

Karin BR, **Das I**, **Jackman TR**, **Bauer AM. 2017.** Ancient divergence time estimates in *Eutropis rugifera* support the existence of Pleistocene barriers on the exposed Sunda Shelf. *PeerJ* **5:** e3762.

Kassambara A, **Mundt F. 2017.** *Package 'Factoextra'. Extract and visualize the results of multivariate data analyses, 76.* Available at: https://cran.r-project.org/web/packages/factoextra/factoextra.pdf (accessed January, 2020).

Katoh K, **Standley DM. 2013.** MAFFT multiple sequence alignment software version **7:** improvements in performance and usability. *Molecular Biology and Evolution* **30:** 772–780.

Kingsley EP, **Kozak KM**, **Pfeifer SP**, **Yang DS**, **Hoekstra HE. 2017.** The ultimate and proximate mechanisms driving the evolution of long tails in forest deer mice. *Evolution* **71:** 261–273.

Kumar S, **Stecher G**, **Tamura, K. 2016.** MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution* **33:** 1870–1874.

Leonard JA, **Den Tex RJ**, **Hawkins MT**, **Muñoz-Fuentes V**, **Thorington R**, **Maldonado JE. 2015.** Phylogeography of vertebrates on the Sunda Shelf: a multi-species comparison. *Journal of Biogeography* **42:** 871–879.

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv preprint arXiv:1303.3997.

Li H, **Handsaker B**, **Wysoker A**, **Fennell T**, **Ruan J**, **Homer N**, **Marth G**, **Abecasis G**, **Durbin R. 2009.** The sequence alignment/map format and SAMtools. *Bioinformatics* **25:** 2078–2079.

Lim HC, **Shakya SB**, **Harvey MG**, **Moyle RG**, **Fleischer RC**, **Braun MJ**, **Sheldon FH. 2020.** Opening the door to greater phylogeographic inference in Southeast Asia: Comparative genomic study of five codistributed rainforest bird species using target capture and historical DNA. *Ecology and Evolution* **10:** 3222–3247.

Martin RD. 1968. Reproduction and ontogeny in treeshrews (*Tupaia belangeri*), with reference to their general behaviour and taxonomic relationships. *Zeitschrift für Tierpsychologie* **25:** 409–495.

Martins RF, **Schmidt A**, **Lenz D**, **Wilting A**, **Fickel J. 2019.** Human-mediated introduction of introgressed deer across Wallace's line. *Ecology and Evolution* **8:** 1465–1479.

Mason VC, **Li G**, **Minx P**, **Schmitz J**, **Churakov G**, **Doronina L**, **Melin AD**, **Dominy NJ**, **Lim NT**, **Springer MS**, **Wilson RK. 2016.** Genomic analysis reveals hidden biodiversity within colugos, the sister group to primates. *Science Advances* **2:** e1600633.

Mason VC, **Helgen KM**, **Murphy WJ. 2019.** Comparative phylogeography of forest-dependent mammals reveals Paleo-forest corridors throughout Sundaland. *Journal of Heredity* **110:** 158–172.

Mason NA, **Fletcher NK**, **Gill BA**, **Funk WC**, **Zamudio KR. 2020.** Coalescent-based species delimitation is sensitive to geographic sampling and isolation by distance. *Systematics and Biodiversity* **18:** 269–280.

McCormack JE, **Heled J**, **Delaney KS**, **Peterson AT**, **Knowles LL. 2011.** Calibrating divergence times on species trees versus gene trees: implications for speciation history of *Aphelocoma* jays. *Evolution* **65:** 184–202.

Medway LGH. 1977. Mammals of Borneo: field keys and an annotated checklist. *Monographs of the Malaysian Branch of the Royal Asiatic Society* **7:** 1–172.

Mein P, **Pickford M. 2006.** Late Miocene micromammals from the Lukeino Formation (6.1 to 5.8 Ma), Kenya. *Bulletin Mensuel de la Société Linnéenne de Lyon* **75:** 183–223.

Merckx VS, **Hendriks KP**, **Beentjes KK**, **Mennes CB**, **Becking LE**, **Peijnenburg KT**, **Afendy A**, **Arumugam N**, **De Boer H**, **Biun A**, **Buang MM. 2015.** Evolution of endemism on a young tropical mountain. *Nature* **524:** 347–350.

Miller GS, **Hollister N. 1921.** Twenty new mammals collected by HC Raven in Celebes. *Proceedings of the Biological Society of Washington* **34:** 93–104.

Moen D, **Morlon H. 2014.** Why does diversification slow down?. *Trends in Ecology & Evolution* **29:** 190–197.

Murphy WJ, **Eizirik E**, **Johnson WE**, **Zhang YP**, **Ryder OA**, **O'Brien SJ. 2001.** Molecular phylogenetics and the origins of placental mammals. Placental mammal diversification and the Cretaceous–Tertiary boundary. *Nature* **409:** 614–618.

Musser GG. 1986. Sundaic Rattus: definitions of Rattus baluensis and Rattus korinchi. *American Museum Novitates* **2862:** 1–24.

Musser GG, **Durden LA**, **Holden ME**, **Light JE. 2010.** Systematic review of endemic Sulawesi squirrels (Rodentia, Sciuridae), with descriptions of new species of associated sucking lice (Insecta, Anoplura), and phylogenetic and zoogeographic assessments of sciurid lice. *Bulletin of the American Museum of Natural History* **2010:** 1–260.

Nations JA, **Heaney LR**, **Demos TC**, **Achmadi AS**, **Rowe KC**, **Esselstyn JA. 2019.** A simple skeletal measurement effectively predicts climbing behaviour in a diverse clade of small mammals. *Biological Journal of the Linnean Society* **128:** 323–336.

Nguyen LT, **Schmidt HA**, **Von Haeseler A**, **Minh BQ. 2015.** IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* **32:** 268–274.

O'Connell KA, **Smart U**, **Smith EN**, **Hamidy A**, **Kurniawan N**, **Fujita MK. 2018.** Within-island diversification underlies parachuting frog (*Rhacophorus*) species accumulation on the Sunda Shelf. *Journal of Biogeography* **45:** 929–940.

Ohtani M, **Kondo T**, **Tani N**, **Ueno S**, **Lee LS**, **Ng KK**, **Muhammad N**, **Finkeldey R**, **Na'iem M**, **Indrioko S**, **Kamiya K. 2013.** Nuclear and chloroplast DNA phylogeography reveals Pleistocene divergence and subsequent secondary contact of two genetic lineages of the tropical rainforest tree species *Shorea leprosula* (Dipterocarpaceae) in South-East Asia. *Molecular Ecology* **22:** 2264–2279.

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.

Padial JM, **Miralles A**, **De la Riva I**, **Vences M. 2010.** The integrative future of taxonomy. *Frontiers in Zoology* **7:** 1–14.

Paradis E, **Schliep K. 2019.** Ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35:** 526–528.

Parham JF, **Donoghue PC**, **Bell CJ**, **Calway TD**, **Head JJ**, **Holroyd PA**, **Inoue JG**, **Irmis RB**, **Joyce WG**, **Ksepka DT**, **Patané JS. 2012.** Best practices for justifying fossil calibrations. *Systematic Biology* **61:** 346–359.

Parker LD, **Hawkins MT**, **Camacho-Sanchez M**, **Campana MG**, **West-Roberts JA**, **Wilbert TR**, **Lim HC**, **Rockwood LL**, **Leonard JA**, **Maldonado JE. 2020.** Little genetic structure in a Bornean endemic small mammal across a steep ecological gradient. *Molecular Ecology* **29:** 4074–4090.

Patterson BD. 2002. On the continuing need for scientific collecting of mammals. *Mastozoología Neotropical* **9:** 253–262.

Pinya S, **Bover P**, **Jurado-Rivera JA**, **Trenado S**, **Parpal L**, **Férriz I**, **Talavera A**, **Hinckley A**, **Pons J**, **López-Fuster MJ**, **Alcover JA. 2018.** Recent island colonization by an introduced shrew in the western Mediterranean. *Hystrix* **29:** 232–235.

R Core Team. **2018.** *R: a language and environment for statistical computing*. Vienna: R Core Team.

Rambaut A, **Drummond AJ**, **Xie D**, **Baele G**, **Suchard MA. 2018.** Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology* **67:** 901.

Reumer JWF. 1989. Speciation and evolution in the Soricidae (Mammalia: Insectivora) in relation with the paleoclimate. *Revue Suisse de Zoologie* **96:** 81–90.

Riddle BR, **Ladle RJ**, **Lourie SA**, **Whittaker RJ. 2011.** Basic biogeography: estimating biodiversity and mapping nature. In: Ladle RJ, Whittaker RJ, ed. *Conservation biogeography*. Oxford: Wiley, 45–92.

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.

Robinson HC, **Kloss CB. 1918.** Results of an expedition to Korinchi Peak, Sumatra, 1. Mammals. *Journal of the Federated Malay States Museum* **8:** 1–80.

Roos C, **Helgen KM**, **Portela-Miguez R**, **Thant NML**, **Lwin N**, **Lin AK**, **Lin A**, **Yi KM**, **Soe P**, **Hein ZM**, **Myint MNN**, **Ahmed T**, **Chetry D**, **Urh M**, **Veatch EG**, **Duncan N**, **Kamminga P**, **Chua MAH**, **Yao L**, **Matauschek C**, **Meyer D**, **Liu ZJ**, **Li M**, **Nadler T**, **Fan PF**, **Quyet LK**, **Hofreiter M**, **Zinner D**, **Momberg F. 2020.** Mitogenomic phylogeny of the Asian colobine genus *Trachypithecus* with special focus on *Trachypithecus phayrei* (Blyth, 1847) and description of a new species. *Zoological Research* **41:** 656–669.

Rowe KC, **Achmadi AS**, **Fabre PH**, **Schenk JJ**, **Steppan SJ**, **Esselstyn JA. 2019.** Oceanic islands of Wallacea as a source for dispersal and diversification of murine rodents. *Journal of Biogeography* **46:** 2752–2768.

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, **Vogel P. 1995.** Chromosomal evolution and zoogeographic origin of Southeast Asian shrews (genus *Crocidura*). *Experientia* **51:** 174–178.

Ruedi M, **Maddalena T**, **Yong HS**, **Vogel P. 1990.** The *Crocidura fuliginosa* species complex (Mammalia: Insectivora) in Peninsular Malaysia: biological, karyological and genetical evidence. *Biochemical Systematics and Ecology* **18:** 573–581.

Ruedi M, **Auberson M.**, **Savolainen V. 1998.** Biogeography of Sulawesian shrews: testing for their origin with a parametric bootstrap on molecular data. *Molecular Phylogenetics and Evolution* **9:** 567–571.

Sarker GC, **Wostl E**, **Thammachoti P**, **Sidik I**, **Hamidy A**, **Kurniawan N**, **Smith EN. 2019.** New species, diversity, systematics, and conservation assessment of the puppet toads of Sumatra (Anura: Bufonidae: *Sigalegalephrynus*). *Zootaxa* **4679:** 365–391.

Shaney KJ, **Maldonado J**, **Smart U**, **Thammachoti P**, **Fujita M**, **Hamidy A**, **Kurniawan N**, **Harvey MB**, **Smith EN. 2020.** Phylogeography of montane dragons could shed light on the history of forests and diversification processes on Sumatra. *Molecular Phylogenetics and Evolution* **149:** 106840.

Sheldon FH, **Lim HC**, **Moyle RG. 2015.** Return to the Malay Archipelago: the biogeography of Sundaic rainforest birds. *Journal of Ornithology* **156:** 91–113.

Sikes RS. 2016. Animal Care and Use Committee of the American Society of Mammalogists. 2016 guidelines of the American Society of Mammalogists for the use of wild mammals in research and education. *Journal of Mammalogy* **97:** 663–688.

Slik JWF, **Aiba SI**, **Bastian M**, **Brearley FQ**, **Cannon CH**, **Eichhorn KAO**, **Fredriksson G**, **Kartawinata K**, **Laumonier Y**, **Mansor A**, **Zweifel N. 2011.** Soils on exposed Sunda Shelf shaped biogeographic patterns in the equatorial forests of Southeast Asia. *Proceedings of the National Academy of Sciences of the USA* **108**: 12343–12347.

Spöri Y, **Flot JF. 2020.** HaplowebMaker and CoMa: two web tools to delimit species using haplowebs and conspecificity matrices. *Methods in Ecology and Evolution* **11:** 1434–1438

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, **Knowles LL. 2017.** Multispecies coalescent delimits structure, not species. *Proceedings of the National Academy of Sciences of the USA* **114:** 1607–1612.

Taylor JM, **Smith SC**, **Calaby JH. 1985.** Altitudinal distribution and body size among New Guinean *Rattus* (Rodentia: Muridae). *Journal of Mammalogy* **66:** 353–358.

Thomas O. 1898. Descriptions of new Bornean and Sumatran mammals. *Annals and Magazine of Natural History* **7:** 247–251.

Van den Bergh GD, **Li B**, **Brumm A**, **Grün R**, **Yurnaldi D**, **Moore MW**, **Kurniawan I**, **Setiawan R**, **Aziz F**, **Roberts RG**, **Suyono, Storey M**, **Setiabudi E**, **Morwood MJ. 2016.** Earliest hominin occupation of Sulawesi, Indonesia. *Nature* **529:** 208–211.

Wallace AR. 1869. *The Malay Archipelago, the land of the orang-utan and the bird of paradise: a narrative of travel with studies of man and nature.* London: Macmillan.

Whittaker RJ, **Fernández-Palacios JM**, **Matthews TJ**, **Borregaard MK**, **Triantis KA. 2017.** Island biogeography: taking the long view of nature's laboratories. *Science* **357:** eaam8326.

William-Dee JA 2019. *Phylogeography of Bornean shrew (family Soricidae: Crocidura foetida) inferred from Cytochrome b gene and species distribution modeling of three shrews from Malaysia using maximum entropy approach.* M.Sc., Unpublished master thesis, Universiti Malaysia Sarawak.

Woodruff DS, **Turner LM. 2009.** The Indochinese-Sundaic zoogeographic transition: a description and analysis of terrestrial mammal species distributions. *Journal of Biogeography* **36:** 803–821.

Xia X. 2018. DAMBE7: new and improved tools for data analysis in molecular biology and evolution. *Molecular Biology and Evolution* **35:** 1550–1552.

Yang L, **Zhang H**, **Zhang C**, **Wu J**, **Wang Z**, **Li C**, **Zhang B. 2020.** A new species of the genus *Crocidura* (Mammalia: Eulipotyphla: Soricidae) from Mount Huang, China. *Zoological Systematics* **45:** 1–14.

Yuan SL, **Jiang XL**, **Li ZJ**, **He K**, **Harada M**, **Oshida T**, **Lin LK. 2013.** A mitochondrial phylogeny and biogeographical scenario for Asiatic water shrews of the genus *Chimarrogale*: implications for taxonomy and low-latitude migration routes. *PLoS One* **8:** e77156.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's web-site.

Table S1. Sample information molecular analyses.
Table S2. Primers used in this study and PCR conditions.
Table S3. Sample information morphological analyses.
Figure S1. Cytochrome *b* maximum-likelihood consensus tree of Asian *Crocidura*.
Figure S2. Species tree of Sundaland and Sulawesi *Crocidura* as estimated from six nuclear loci using *BEAST.
Figure S3. Haplowebs for twelve polymorphic nuclear genes studied in the *Crocidura neglecta* complex.
Figure S4. Conspecificity matrix obtained using CoMa, integrating the results from combining the delimitations produced by the 12 nuclear loci of Figure S5.
Figure S5. Haplowebs for twelve of the thirteen polymorphic nuclear genes studied in the Sabah-Philippines-Sulawesi clade.
Figure S6. Haplowebs for two nuclear genes studied in the Sabah-Philippines-Sulawesi clade and including historic samples of *Crocidura baluensis* from Kinabalu.
Figure S7. Conspecificity matrix obtained using CoMa, integrating the results from the 12 nuclear loci of Figure S3.
Figure S8. Picture highlighting hindfoot differences along elevation in the "*Crocidura baluensis-C. foetida s. l.* from Sabah" lineages.
File S1. File including supplementary Figures S1–S8 (phylogenetic trees, haplowebs, conspecificity matrices and morphological variation in the *Crocidura baluensis–C. foetida* s.l. from Sabah' lineage.
File S2. Zip file containing complete DNA sequence alignments.
File S3. Initial mitochondrial genome partitions; best-fit partition scheme mitochondrial genome; IQTREE runs results; PCAs results.