

RESEARCH PAPER

Isolation by marine barriers and climate explain areas of endemism in an island rodent

Thomas C. Giarla¹  | Sean P. Maher² | Anang S. Achmadi³ | Meagan K. Moore⁴ | Mark T. Swanson^{4,5} | Kevin C. Rowe^{6,7} | Jacob A. Esselstyn^{4,5}

¹Department of Biology, Siena College, Loudonville, New York

²Department of Biology, Missouri State University, Springfield, Missouri

³Museum Zoologicum Bogoriense, Research Center for Biology, Cibinong, Indonesia

⁴Museum of Natural Science, Louisiana State University, Baton Rouge, Louisiana

⁵Department of Biological Sciences, Louisiana State University, Baton Rouge, Louisiana

⁶Sciences Department, Museums Victoria, Melbourne, VIC, Australia

⁷BioSciences, University of Melbourne, Melbourne, VIC, Australia

Correspondence

Thomas C. Giarla, Siena College, 515 Loudon Rd, Loudonville, NY 12211.
Email: tgiarla@siena.edu

Funding information

NIH, Grant/Award Number: 5R25GM102765-02; NSF, Grant/Award Number: DEB-1343517, DEB-1441634, OISE-096586; National Geographic Society, Grant/Award Number: 9025-11; James and Carol Patton; RISTEK

Editor: Krystal Tolley

Abstract

Aim: The Indonesian island of Sulawesi is home to thousands of endemic terrestrial animal species. Phylogeographical studies of some of these taxa revealed concordant distribution patterns (areas of endemism; AoEs), suggestive of a shared evolutionary or ecological mechanism driving divergence among various terrestrial taxa. Generally, AoEs have been attributed to Pleistocene marine incursions, geological fault zones and Sulawesi's history as an archipelago. We test, for the first time, the hypothesis that population divergences are associated with unsuitable climate spaces at the boundaries between these areas.

Location: Sulawesi, Indonesia.

Taxon: *Maxomys musschenbroekii*, a common murid rodent endemic to Sulawesi.

Methods: We sequenced a mitochondrial gene, four nuclear protein-coding genes and hundreds of ultraconserved elements from individuals sampled across the island, which we used to build intraspecific phylogenies and identify population structure. To test climate-related hypotheses, we used multiple regression to assess the extent to which genetic distances, over-land geographical distances, environmental distances (derived from ecological niche model resistance surfaces) and areas of endemism are correlated.

Results: In all three genetic datasets, we inferred phylogeographical structure that is geographically concordant with previously defined AoEs. Ecological niche models showed a similar geography of suitability across the island since the Last Glacial Maximum, with varying levels of unsuitable climate space between AoEs. Genetic distances were more strongly correlated with environmental distances than geographical distances and interspecific divergences arose just in the past 1.5 Myr.

Main Conclusions: Population divergences within *M. musschenbroekii* are consistent with AoE definitions from other taxa. Stable areas of unsuitable habitat near AoE boundaries probably supplemented marine incursions in generating population structure within *M. musschenbroekii* and other co-distributed species.

KEYWORDS

areas of endemism, ecological niche models, isolation by environment, *Maxomys musschenbroekii*, Muridae, phylogeography, Sulawesi, ultraconserved elements

1 | INTRODUCTION

Island ecosystems provide discrete, practical venues for examining processes of biological diversification (Gillespie, 2004). Whereas much of our understanding of insular diversification is derived from investigations of between-island isolation (e.g., Grant & Grant, 2008; Lack, 1947), recent work has begun to acknowledge the importance of processes that occur within large islands (e.g., Demos et al., 2016; Esselstyn, Maharadatunkamsi, Achmadi, Siler, & Evans, 2013; Justiniano et al., 2014; Losos, Jackman, Larson, Queiroz, & Rodriguez-Schettino, 1998; Setiadi et al., 2011; Toussaint et al., 2014). In some cases, radiations of species have evolved *in situ* over just a few million years (e.g., *Apomys* and relatives on Luzon in the Philippines; Heaney et al., 2011; Justiniano et al., 2014; Heaney et al., 2016). Nonetheless, many large tropical islands—especially those of the Wallacean biogeographical zone—remain relatively unexplored in terms of their potential for the discovery of biogeographical mechanisms.

The island of Sulawesi, in Indonesia, is the centrepiece of Wallacea (Figure 1a) and is noted for consistent patterns of geographical structure among many of its terrestrial animals (Evans, Supriatna, Andayani, Setiadi et al., 2003; Lohman et al., 2011). Fooden (1969), in a study of the morphological variation among Sulawesi's seven parapatrically distributed macaque species, identified hybrid zones at species' range edges and speculated that the island's geological history might explain the location of these areas. The geographical structure of morphological and genetic variation in these macaques is now relatively well studied (Bynum, Bynum, & Supriatna, 1997; Ciani, Stanyon, Scheffrahn, & Sampurno, 1989; Evans, Morales, Supriatna, & Melnick, 2008; Evans, Supriatna, Andayani, & Melnick, 2003), and their hybrid zones, together with concordant phylogeographical breaks in two groups of amphibians (Evans, Brown et al., 2003; Evans, Supriatna, Andayani, Setiadi et al., 2003), were used to define seven areas of endemism (AoE) on Sulawesi (Figure 1b). Studies of other animal groups also have found extensive, geographically partitioned diversity, much of which is concordant with at least some of the AoE boundaries (see debate in Bridle, Pedro, & Butlin, 2004; Evans, Cannatella, & Melnick, 2004; Evans, McGuire, Brown, Andayani, & Supriatna, 2008; Eldridge, Achmadi, Giarla, Rowe, & Esselstyn, 2018).

Concordant patterns of genetic diversity among taxa with disparate natural history traits, as seen in Sulawesi animals, suggest a shared mechanism of diversification (Arbogast & Kenagy, 2001; Evans, Supriatna, Andayani, Setiadi et al., 2003). On Sulawesi, these concordant patterns have been linked to the island's geological history and geographical features (Driller et al., 2015; Eldridge et al., 2018; Evans, Brown et al., 2003; Evans, Supriatna, Andayani, Setiadi et al., 2003; Merker, Driller, Perwitasari-Farajallah, Pamungkas, & Zischler, 2009; Setiadi et al., 2011; Stelbrink, Albrecht, Hall, & von Rintelen, 2012). This large (174,600 km²), mountainous (six peaks >3,000 m), "K"-shaped island has a complex palaeogeographical history. Nugraha and Hall (2018) synthesized evidence from the

geological record to reconstruct the palaeogeography of Sulawesi and its precursors over the last 20 Ma. They hypothesized that, 5 Ma, the island's current components were mostly below water. At that time, a cluster of six smaller islands were isolated by shallow seas <200 m deep. Between 5 and 1 Ma, continued tectonic activity and sustained uplift united most of the smaller islands and led to the formation of mountains in several areas. At 1 Ma, only the present-day Southwest peninsula was cut off from the rest of the island at the intermittently inundated Tempe Depression. That shallow marine zone is reconstructed as having been dry during glacial maxima and inundated during glacial minima (Gremmen, 1990; Nugraha & Hall, 2018). In its present form, Sulawesi's land area is dominated by peninsulas and numerous mountainous regions. Fault lines across the island demarcate geological suture zones, regions that have experienced high tectonic activity and mountain uplift in the past several million years (Bellier et al., 2006; Hall, 2002; Nugraha & Hall, 2018). Many of the resulting mountain ranges are isolated by intervening lowlands, which may represent a habitat barrier to highland species (Eldridge et al., 2018).

We evaluate three hypotheses to explain the consistent patterns of geographical isolation among populations of Sulawesi's terrestrial taxa. First, genetic subdivisions were caused by null processes such as isolation-by-distance or peninsular effects that yield phylogeographical breaks without discrete dispersal barriers (Irwin, 2002). Second, we follow previous studies and propose that populations diverged due to physical barriers to dispersal. Barriers may have existed on or between isolated palaeo-islands or after island coalescence via a number of landscape features (e.g., marine incursions of lowland areas, formation of rivers, or uplift of mountain ridges). Our third hypothesis, which has not been considered previously, is that climatic variation across Sulawesi drove concordant patterns of lineage diversification by creating climatic barriers to dispersal at AoE boundaries. If organisms are unable to disperse across unsuitable climatic spaces (e.g., lowlands) or they adapt to a distinct and geographically restricted climatic space, climatic variation could generate concordant phylogeographical patterns. This last hypothesis is primarily relevant for the time period during which most of the present-day island was above water and contiguous (i.e., 1 Ma to the present; Nugraha & Hall, 2018).

Maxomys musschenbroekii, a small, omnivorous murid rodent that lives in primary and secondary forest, is an excellent candidate for exploring the AoE paradigm on Sulawesi and testing the hypotheses discussed above. The species is one of the most common and widespread terrestrial small mammals on the island, with an elevational range from near sea level to well over 2,000 m (museum records). Although it is known from most forested sites on the island that have been surveyed for small mammals, it has never been the focus of phylogeographical analyses. Achmadi, Esselstyn, Rowe, Maryanto, and Abdullah (2013) estimated a phylogeny for 13 of the 18 recognized species of *Maxomys* and included individuals of *M. musschenbroekii* from six sites on the island, mostly sampled from the West-Central AoE. Achmadi et al. (2013) found three divergent

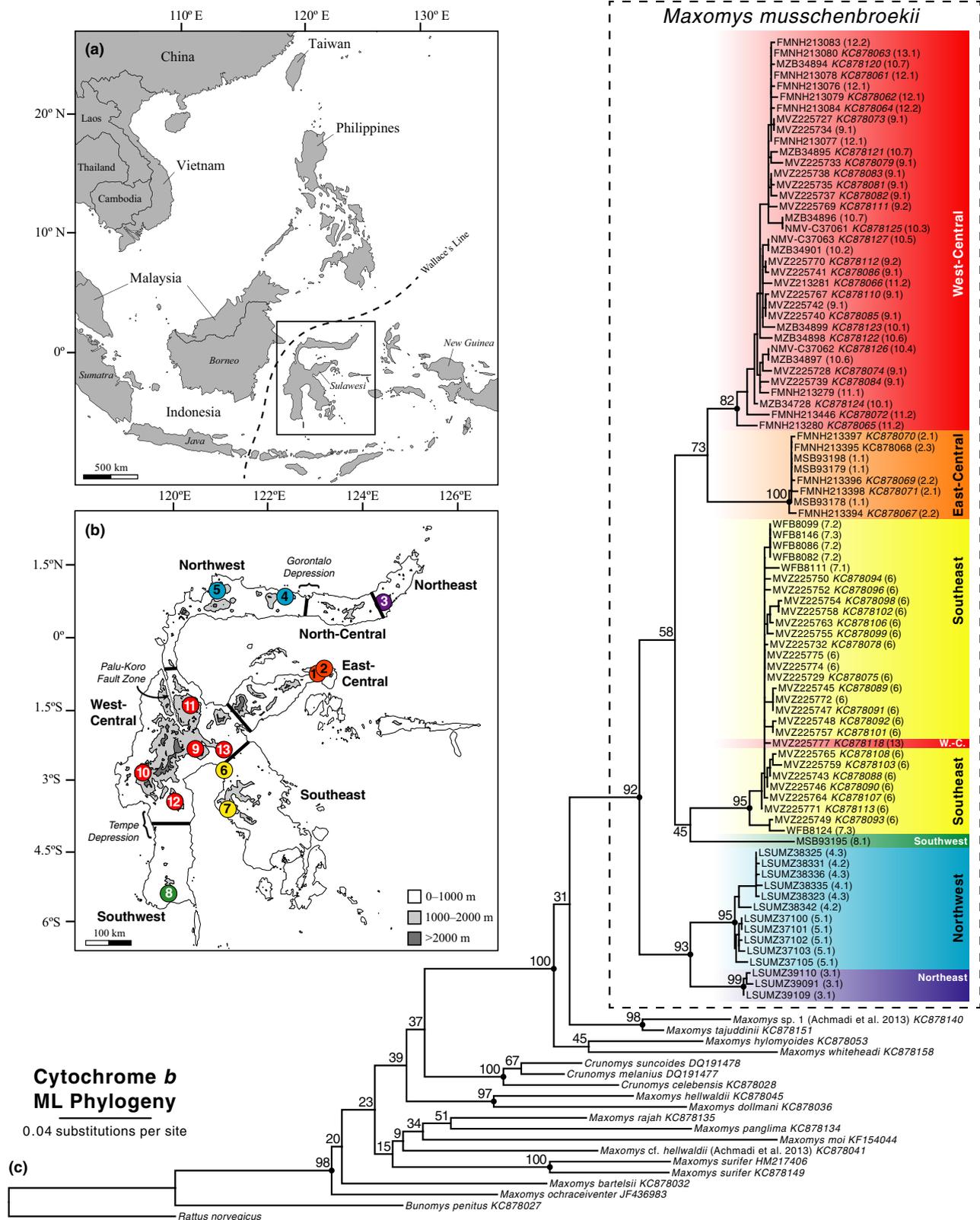


FIGURE 1 (a) Map of Wallacea and surrounding region. (b) Map of Sulawesi, showing clusters of *Maxomys musschenbroekii* collection sites, elevation and notable landscape features. Only sites with genetic data are shown. When collecting sites were geographically clustered (<10 km), a single symbol is portrayed on the map. Dark black lines denote area of endemism (AoE) boundaries (Evans, Supriatna, Andayani, Setiadi et al., 2003). (c) Maximum-likelihood (ML) phylogeny of cytochrome *b* sequences. Numbers near nodes denote percent bootstrap support, with filled dark circles indicating support $\geq 75\%$. Support values at nodes within coloured clades not shown. Accession numbers for sequences downloaded from GenBank are provided in italics after the museum voucher number. Museum acronyms are defined in the text. Collection site numbers are provided in parentheses and explained in Table S1.1

mitochondrial lineages within the species (maximum uncorrected p -distance = 0.07 for cytochrome b); each mitochondrial clade was restricted to a single AoE, hinting that the same process that drove divergence in other terrestrial Sulawesi taxa might also be at work in *M. musschenbroekii*. In this study, we explore the relevance of the AoE paradigm to *M. musschenbroekii* with greatly expanded samples of sites and genetic loci. Our primary goals were to test whether differentiation among populations (both genetic and morphological) is concordant with AoE boundaries, and if so, are the breaks in these data best explained by geographical distance (reflecting isolation by distance), past and present physical barriers, or areas of climatic suitability over the Quaternary.

2 | MATERIALS AND METHODS

2.1 | Specimens and morphology

Our study includes voucher specimens (including preserved tissues, when available) from the Field Museum of Natural History (FMNH); Louisiana State University Museum of Natural Science (LSUMZ); Museum Zoologicum Bogoriense (MZB); University of California at Berkeley Museum of Vertebrate Zoology (MVZ); University of California at Davis Museum of Wildlife and Fish Biology (WFB); and University of New Mexico Museum of Southwestern Biology (MSB). These specimens (Table S1.1 in Supporting Information) are from 52 collection sites across Sulawesi. Sites <10 km apart from one another are grouped into 22 clusters (clusters 1–13, which include specimens with genetic data, are mapped in Figure 1b; the remaining clusters are not mapped but are represented in Table S1.1). We gathered external measurements (total length, tail length, hind foot length, ear length and mass) for 195 adult specimens from the field notes of collectors. To qualitatively determine whether the genetic divergences, we observed between AoE populations (see below) are associated with phenotypic differences, we conducted a principal components analysis of the correlation matrix of all five measurements and plotted various bivariate combinations of raw measurements.

2.2 | Sanger DNA sequencing

We had access to tissue samples of *Maxomys musschenbroekii* individuals from six of the seven AoEs defined on Sulawesi (Evans, Supriatna, Andayani, Setiadi et al., 2003), and we selected 70 of these for DNA sequencing (Table S1.1). We used PCR and published primers (Table S1.2) to amplify the first 801 bp of the mitochondrial gene cytochrome b (CYTB) and exon segments of four protein-coding nuclear genes (PCNGs: the X-linked ATPase copper transporting alpha [ATP7A] gene, the cannabinoid receptor 1 gene [CNR1], the growth hormone receptor gene [GHR] and the interphotoreceptor retinoid-binding protein gene [IRBP]). We used GoTaq Green Master Mix (Promega) for amplification and followed the recommended thermal cycler protocol for that reagent, using primer-specific annealing temperatures (Table S1.2). Purified PCR products were

sequenced at the Cornell University Institute of Biotechnology (Ithaca, NY) on an ABI 3730xl using Big Dye Terminator v3.1 chemistry (ThermoFisher). In GENEIOUS R9 (BioMatters), we edited chromatograms, called heterozygotes, and aligned sequences via MUSCLE (Edgar, 2004). To resolve heterozygotes in the PCNGs, we algorithmically phased the sequences in PHASE 2.1.1 (Stephens, Smith, & Donnelly, 2001) with a probability threshold of 0.7. Heterozygote bases that were unable to be phased under those parameters were left as their corresponding IUPAC ambiguity code. All sequences are available on GenBank with accession codes MH068879–MH069205 (Table S1.1).

2.3 | Mitochondrial phylogeny

Our first analytical goal was to build a densely sampled mitochondrial phylogeny for *Maxomys musschenbroekii* and other related species. To assess the monophyly of *M. musschenbroekii*, we compiled additional CYTB sequences from other *Maxomys* species with data available on GenBank. This set included three *Crnomys* species—which Achmadi et al. (2013) identified as being nested within *Maxomys*—along with two outgroup species from closely related genera (*Bunomys penitus* and *Rattus norvegicus*). We also included GenBank CYTB sequences from other *M. musschenbroekii* specimens for which we did not have fresh tissues available for this study. We aligned our new CYTB sequences and the sequences from GenBank in Geneious using MUSCLE. We assessed the fit of nucleotide substitution models in jMODELTEST 2.1.7 (Darriba, Taboada, Doallo, & Posada, 2012) and applied the best-fitting model according to the Bayesian Information Criterion to a maximum-likelihood (ML) phylogenetic analysis in PHYLML 3.0 (Guindon et al., 2010). We calculated nodal support using 100 bootstrap replicates.

2.4 | Ultraconserved element sequencing and assembly

We selected 11 *Maxomys musschenbroekii* individuals from six AoEs for ultraconserved element (UCE) sequencing and assembly (Table S1.1). UCE sequences have primarily been used to build interspecific phylogenies (Faircloth et al., 2012), but, as long as the variable regions adjacent to UCES are included, they offer numerous advantages over other common markers (e.g., RAD-Seq) used in phylogeographical studies (Harvey, Smith, Glenn, Faircloth, & Brumfield, 2016). These include repeatability and being comparable across more phylogenetically distant taxa. To sequence UCES and the variable flanking regions, we followed the protocols of Giarla and Esselstyn (2015). Briefly, genomic libraries were prepared using a Kapa “on bead” Library Preparation Kit (Kapa Biosystems). We baited the libraries using a UCE MYbaits probe set (MYcroarray) that targets $c.$ 2,500 loci. We sequenced the enriched, pooled libraries on an Illumina HiSeq (150bp, paired-end) at the Georgia Genomics Facility (Athens, GA). Reads were subjected to quality control and trimmed using TRIMMOMATIC 0.32 (Bolger, Lohse, & Usadel, 2014). We used PHYLUCE 1.5 (Faircloth, 2016) to automate the *de novo* assembly of

reads for each sample in TRINITY 2.0.6 (Grabherr et al., 2011). We also used PHYLUCe to identify contigs matching UCE loci and to align those loci via MUSCLE, discarding any alignment that included 3 or fewer individuals. Raw Illumina reads and UCE sequences are included in NCBI BioProject PRJNA437989; individual Targeted Locus Study accession numbers for UCEs are provided in Table S1.1.

2.5 | Concatenated UCE phylogeny

No other *Maxomys* species have UCE sequences publicly available for use as an outgroup, so we used the mouse (*Mus musculus*) UCE sequences that Faircloth et al. (2012) extracted bioinformatically from the mouse genome. To minimize the potential effects of missing data on phylogeny estimation, we filtered our UCE loci to only include those with 75% or more of the individuals represented and those with at least one variable site. After the *M. musculus* sequences were aligned to our *Maxomys* sequences, we concatenated all of the loci in Geneious. We inferred a maximum-likelihood phylogeny with bootstrapping using the GTRCAT model in RAxML 8 (Stamatakis, 2014) on the CIPRES Science Gateway (Miller, Pfeiffer, & Schwartz, 2010). We retained the default settings and let the program stop bootstrapping automatically according to the majority-rule criterion (autoMRE).

2.6 | Population structure in BPP

We assigned *Maxomys musschenbroekii* individuals to six putative “species” that correspond to six well-supported mitochondrial clades, each nearly entirely restricted to a single AoE (see Results). We tested for population structure in Bayesian Phylogenetics and Phylogeography (BPP) 3.3 (Yang, 2015; Yang & Rannala, 2010, 2014) using two datasets: (1) a geographically widespread set of 70 *M. musschenbroekii* individuals with sequences from four PCNGs and (2) a subset of 11 individuals (1 or 2 from each AoE) sequenced for hundreds of UCEs. For all analyses, we used BPP’s A11 mode to avoid specifying a guide tree in advance. For the PCNG dataset, we conducted four replicates each for four different combinations of priors on branch lengths and effective population sizes (τ and θ , respectively; Table 1). For the UCE dataset, we observed poor

TABLE 1 Prior schemes (PS) used in guide-tree-based and pairwise BPP analyses of *Maxomys musschenbroekii* populations in Sulawesi. Prior distributions on τ represent two relative divergence depths (deep and shallow) and on θ represent two relative mutation-rate-scaled effective population sizes (large and small)

Prior scheme (PS)	Divergence depth	Effective pop. size	Gamma distribution for prior
1	Deep	Large	$\theta \sim \Gamma(1, 10)$ & $\tau \sim \Gamma(1, 10)$
2	Shallow	Large	$\theta \sim \Gamma(1, 10)$ & $\tau \sim \Gamma(2, 2000)$
3	Deep	Small	$\theta \sim \Gamma(2, 2000)$ & $\tau \sim \Gamma(1, 10)$
4	Shallow	Small	$\theta \sim \Gamma(2, 2000)$ & $\tau \sim \Gamma(2, 2000)$

mixing on trial runs that included all of the loci at once (i.e., replicated runs on the same dataset produced markedly different results). Therefore, we constructed 10 smaller UCE datasets, each comprising 500 randomly selected UCE loci. For each of those 10 datasets, we ran BPP four times, once for each of the four prior schemes, and results were more consistent. For all BPP analyses, we started the MCMC with a 25,000 generation burn-in period where no parameter values were recorded. This was followed by 500,000 MCMC generations, which we sampled every 10 generations.

2.7 | Time-calibrated species tree

Either five or six mitochondrial lineages were consistently delimited as “species” according to BPP analyses of nuclear loci (both UCEs and PCNGs; see Results). We inferred two species trees, one with all six mitochondrial lineages designated as distinct species, the other with five (where the West-Central and Southeast mitochondrial lineages were grouped as one, following the results of the BPP analysis of UCE loci). We inferred species trees using the STARBEAST2 algorithm (Ogilvie, Bouckaert, & Drummond, 2017) in BEAST 2.4.5 (Bouckaert et al., 2014) and a dataset that included CYTB and the four phased PCNGs. *Rattus norvegicus* served as an outgroup. To time-calibrate the species trees, we assigned a uniform prior from 0.01 to 0.03 substitutions per site per lineage per million years on the clock rate for CYTB, a conservative range that accommodates the higher rates of nucleotide substitution observed for murine rodents relative to other mammals (Arbogast, Edwards, Wakeley, Beerli, & Slowinski, 2002; Hardy, González-Cózatl, Arellano, & Rogers, 2013; Wu & Li, 1985). The PCNGs were all assigned uncorrelated lognormal clocks with an exponential prior on the mean rate. The appropriate ploidy settings were applied to the X-linked ATP7A gene and the mitochondrial CYTB gene. We tested alternative nucleotide substitution models for the phased alignments for each locus in JMODELTEST and assigned the best-fitting model to the corresponding locus in BEAST. We ran BEAST for one billion generations, sampling the MCMC chain every 100,000 generations. We discarded the first 10% of trees in the posterior sample as burn-in and produced a maximum clade credibility tree in TREEANNOTATOR 2.4.5 (Bouckaert et al., 2014).

2.8 | Ecological niche models

To calculate environmental distances between sample sites, we first constructed ecological niche models (ENMs) for *Maxomys musschenbroekii* using MAXENT 3.3.3k (Phillips, Anderson, & Schapire, 2006; Phillips & Dudík, 2008) and georeferenced collection sites (Table S1.1). Museum records were downloaded from VertNet, and each record was checked for georeferencing accuracy. To minimize spatial redundancy, the initial 52 unique sites were thinned with a threshold of 10 km using the spThin package (Aiello-Lammens, Boria, Radosavljevic, Vilela, & Anderson, 2015) in R 3.3.1 (R Core Team, 2016), resulting in a set of 22 points. To investigate how environmental distances might have changed over time, we obtained climate

layers at 2.5 arc-minutes from WorldClim.org (1.4, Hijmans, Cameron, Parra, Jones, & Jarvis, 2005) representing the present (1960–1990), the mid-Holocene (~6,000 yr BP) and the Last Glacial Maximum (LGM, ~22,000 yr BP). For the historical datasets, three climate circulation models (CCSM4, MIROC and MPI) were available for both time frames. Climate data were available as bioclimatic variables (Hijmans et al., 2005). We masked the modern data to the current extent of Sulawesi using ArcGIS 10 (ESRI, 2012) and the “raster” package (Hijmans & Etten, 2014) in R. We then calculated correlation coefficients among variables, removing those with high r values (≥ 0.70) to minimize collinearity. MAXENT tends to generate overly complex models under its default settings, which is problematic when transferring models across time or space (Warren & Seifert, 2011). We used ENMEVAL (Muscarella et al., 2014) to identify settings for the best-supported MAXENT models, which were those within 2 AICc units of the top model. The models were then re-run using fivefold cross-validation and projected onto the historical climate datasets. To create consensus maps, we calculated AICc weights based only on the top models, applied the weights to the mean maps from cross-validation, and took the sum for each of the seven scenarios (modern, three mid-Holocene and three LGM). To evaluate potential changes in suitability in the landscape, we calculated similarity scores (Schoener's D and Hellinger's I) between the consensus modern map and each of the consensus hindcast maps using the “dismo” package (Hijmans, Phillips, Leathwick, & Elith, 2015) in R. As a metric of congruence among retrodictions, we calculated the similarity scores for consensus hindcast maps between circulation models.

2.9 | Isolation by environment and by distance

To build a regression model examining the effects of isolation by geographical distance versus isolation by environment (while accounting for AoE assignments), we compiled four matrices: patristic genetic distances, over-land geographical distances, environmental distances and a matrix of AoE assignments. We computed ML-corrected patristic distances in Patristic 1.0 (Fourment & Gibbs, 2006) using the CYTB ML phylogeny with outgroups pruned out. For locations represented in the CYTB ML phylogeny, we used CIRCUITSCAPE (Shah & McRae, 2008) to measure both over-land geographical distances and environmental distances. CIRCUITSCAPE uses raster layers as friction surfaces and a random walk algorithm to estimate how the

landscape might hinder or facilitate movement between pairs or among sets of nodes. For the simple over-land geographical distances, we used a “flat” surface—which did not take environmental suitability into account—and sampled areas as nodes. For the environmental distances, we used consensus maps of our ecological niche models as conductance surfaces and sampled areas as nodes. To accommodate environmental variability over time and the shifting extent of the island due to sea level changes, we estimated both geographical and environmental distance at present, the mid-Holocene and the LGM. Finally, to compute the AoE matrix, we assigned specimens to AoEs using the amphibian and macaque boundaries from Evans, Supriatna, Andayani, Setiadi et al. (2003) and constructed a simple distance matrix in which individuals were designated as belonging to the same (0) or different AoEs (1).

We used multiple matrix regression with randomization in R (MMRR; Wang, 2013; script available from Dryad Digital Repository, <https://doi.org/10.5061/dryad.kt71r>) to quantify contributions of genetic isolation by geographical distance and isolation due to differences in ENM suitability, while accounting for AoE affiliation. Environmental and geographical distance matrices were scaled to allow comparison of coefficients. We used a binomial test to determine whether the proportion of models in which the coefficient associated with ENM-related isolation had a larger magnitude than isolation by random-walk distance was different from random.

3 | RESULTS

3.1 | Phylogenies

The four genetic data matrices we compiled differ in the number of loci, length of the combined alignment, amount of missing data and number of variable sites (Table 2). The best-fitting nucleotide substitution model identified by JMODELTEST for CYTB was GTR+I+ Γ . The phylogeny estimated in PHYML under this model shows *Crnomys* nested within *Maxomys* and strong (92%) bootstrap support for the monophyly of *M. musschenbroekii* (Figure 1c). *Maxomys musschenbroekii* comprises five well-supported (bootstrap support $\geq 82\%$) mitochondrial clades and one divergent sequence (corresponding to *M. musschenbroekii* voucher MSB93195, our only sample from the Southwest AoE). These clades correspond geographically to six of the seven Sulawesi AoEs, with the exception of MVZ225777. That individual, which was collected from the West-Central AoE (cluster

TABLE 2 Summary of genetic data from *Maxomys musschenbroekii* populations in Sulawesi, excluding outgroups

Dataset name	Analysis type	Number of individuals	Number of loci	Total length (bp)	% Missing data	Total number of variable sites
Cytochrome <i>b</i> (CYTB)	BEAST, PhyML	88	1	801	2.2%	172
Protein-coding nuclear genes (PCNGs)	BEAST, BPP	70	4	2,579	8.0%	86
All ultraconserved elements (UCEs)	BPP	11	1,544	1,219,679	50.5%	5,132
75% Complete UCEs	RAxML	11	841	424,216	23.2%	2,669



13 in Figure 1b), is nested within the clade containing all of the specimens from the Southeast AoE (Figure 1c). Relationships among the six different lineages are generally poorly supported, except for the strong support for a sister relationship between the two clades from Sulawesi's northern peninsula (Northwest and Northeast AoEs).

The time-calibrated species tree generated from a combined analysis of CYTB and four PCNGs has a similar topology to the CYTB tree, except the Southeast clade is part of a strongly supported group that also includes the West-Central and East-Central lineage (Figure 2a). When individuals from the West-Central and Southeast AoEs were assumed to be the same species, the resulting topology, relative branch lengths and support values are nearly identical to the 6-taxon species tree (Fig. S1.1). The deepest coalescence within the *M. musschenbroekii* gene tree is estimated to have occurred 1.5 Ma (highest posterior density interval [HPDI]: 0.94–2.09). An ML phylogeny of concatenated UCes has a similar topology as the species tree, except for the position of the West-Central lineage relative to the Southeast and East-Central lineages (Figure 2b). Unlike the CYTB phylogeny, the monophyly of the Southeast AoE and West-Central AoE are only weakly supported.

3.2 | Population structure and morphology

Results from a series of replicated BPP analyses using four PCNGs were generally in agreement regarding population structure among AoE clades, with a minor effect of prior choice on probability of “species” delimitation (summarized in Table 3; full results in Table S1.3). With a posterior probability (PP) threshold of 0.95, prior

Schemes 1 and 2 failed to delimit the Southwest and Northeast mitochondrial clades (PP ranging from 0.78 to 0.87 across the replicates), whereas Prior Schemes 3 and 4 both consistently did (PP \geq 0.98). The other four mitochondrial clades were each delimited with stronger support (PP \geq 0.93) across all prior schemes.

We conducted BPP runs using UCE loci in four sets, one for each prior scheme. Each set included 10 BPP runs using 500 randomly selected loci from the 1,544 available UCes. Posterior probabilities for all delimited species were extracted from the output files (Table S1.4) and averaged across the replicates within each prior scheme set. All six of the AoE clades were strongly supported in individual BPP replicates. For four clades (Northeast, Northwest, Southwest and East-Central), that support was largely consistent across all replicates (Table 3; Fig. S1.2; average PP of species delimitation \geq 0.95, ranging from 0.60 to 1.0). For two clades (West-Central and Southeast), support varied more widely across replicates and prior schemes (Table 3; Fig. S1.2; average PP of species delimitation $<$ 0.95, ranging from 0.12 to 1.0). In replicates where support for West-Central and Southeast was less than 0.95, BPP allocated the remaining support for a combined “species” that groups the adjacent West-Central and Southeast populations together (Table 3; Fig. S1.2; average PP of species delimitation $<$ 0.5, ranging from 0.003 to 0.88).

Bivariate plots of raw external measurements and the first two principal components (Figs. S1.3 and S1.4) showed extensive overlap in the phenotypic variation in each AoE. As such, there is no clear evidence of any AoE harbouring a phenotypically distinctive population, at least with regard to external measurements.

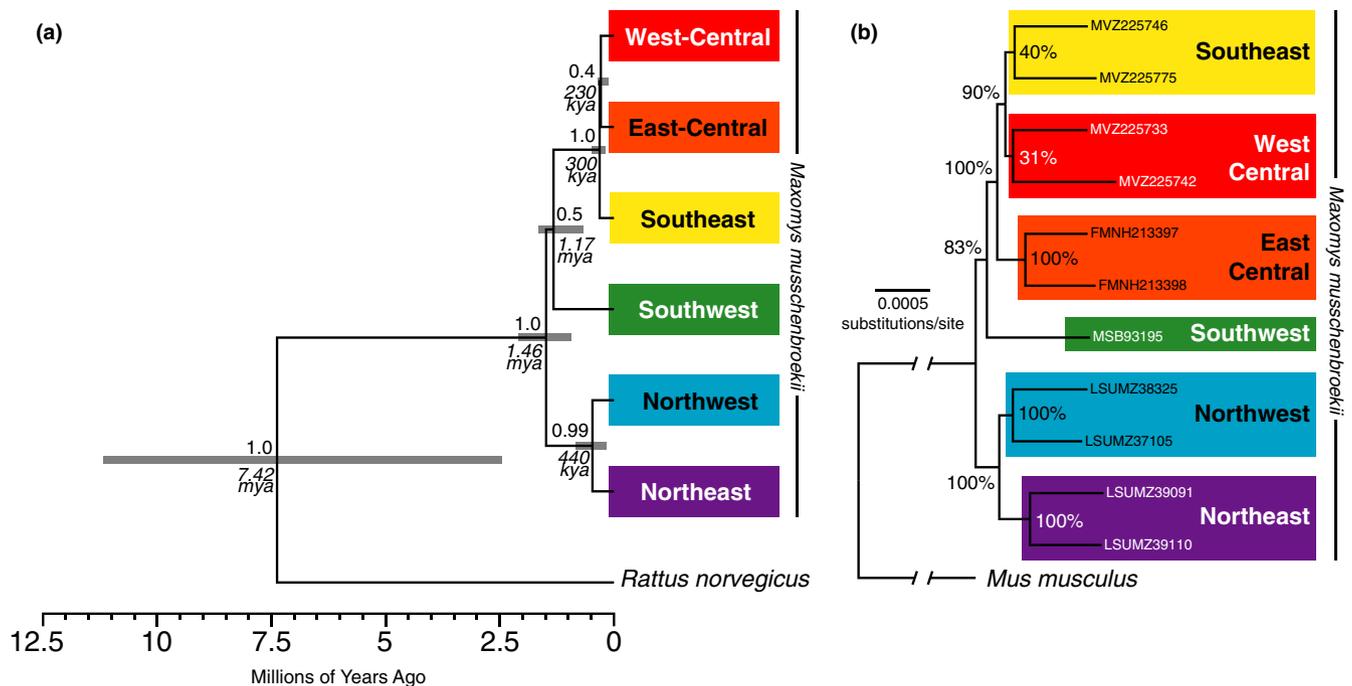


FIGURE 2 (a) Time-calibrated species tree estimated in BEAST using the CYTB and protein-coding nuclear gene dataset for *Maxomys musschenbroekii* in Sulawesi. Numbers above nodes indicate posterior probabilities and grey bars denote 95% highest-posterior density interval for divergence time. (b) ML phylogeny estimated in RAXML using the concatenated 75% complete UCE dataset. Branch lengths to the outgroup are not to scale. Numbers at nodes indicate bootstrap support

TABLE 3 Summary of BPP results from *Maxomys musschenbroekii* populations in Sulawesi for two datasets: 4 protein-coding nuclear genes (PCNGs) and hundreds of ultraconserved element (UCE) loci

Putative AoE Clade ^a	PCNGs ^b					UCEs ^c				
	PS1 average	PS2 average	PS3 average	PS4 average	Overall average	PS1 average	PS2 average	PS3 average	PS4 average	Overall average
West-Central	1.000	1.000	0.991	0.992	0.996	0.770	0.785	0.584	0.584	0.681
Southeast	1.000	1.000	0.991	0.992	0.996	0.778	0.796	0.584	0.584	0.685
East-Central	1.000	1.000	0.991	0.992	0.996	1.000	1.000	0.987	0.954	0.985
Northwest	0.931	0.939	0.998	0.999	0.967	0.993	0.999	0.988	0.988	0.992
Northeast	0.847	0.870	0.992	0.990	0.925	0.993	0.992	0.988	0.988	0.990
Southwest	0.778	0.810	0.981	0.980	0.887	0.981	0.953	0.986	0.943	0.966
West-Central + Southeast	0.000	0.000	0.000	0.000	0.000	0.215	0.200	0.503	0.449	0.342

Note. ^aOther “species” (combinations of populations from different Areas of Endemism [AoE]) received lower support. Full results that include that information are available in Tables S1.4 and S1.5.

^bValues shown are average posterior probabilities (PP) of delimitation from four replicated runs of BPP under each of four different prior schemes (PS; Table 1).

^cValues are average PPs from ten 500-locus subsets under each PS. Only the seven putative “species” that receive posterior probabilities >0.25 on average in either dataset are shown here.

3.3 | Ecological niche models

After examining correlations between bioclimatic variables, we selected eight variables (Bioclim 1, 2, 3, 4, 7, 12, 15 and 19) to use with MAXENT. The most preferred MAXENT model was limited to linear features and used a regularization term of two. There were seven additional models within two AICc units of this model (Table S1.5). The settings from these eight models were used to generate new models with cross-validation for prediction and hindcasting. The resultant maps (Fig. S1.5) were used to generate consensus maps for each scenario (Figure 3). Overall, models had similar mean AUC

scores, and there was little difference between mean training and mean testing values (Table S1.5). The modern ecological niche model illustrates a broad range for *Maxomys musschenbroekii* across Sulawesi, with highly predicted areas mostly limited to higher elevations (Figure 3a). Zones of low-predicted area are present near the borders of all AoEs, and there are extensive areas within each area that are not predicted to be suitable habitat for the species. Hindcasted models at the mid-Holocene (Figure 3b) and LGM (Figure 3c) are somewhat more extensive, illustrating a geographically broader area of potential suitability for this species in the past. Nonetheless, regardless of circulation model, estimates of suitability in the past

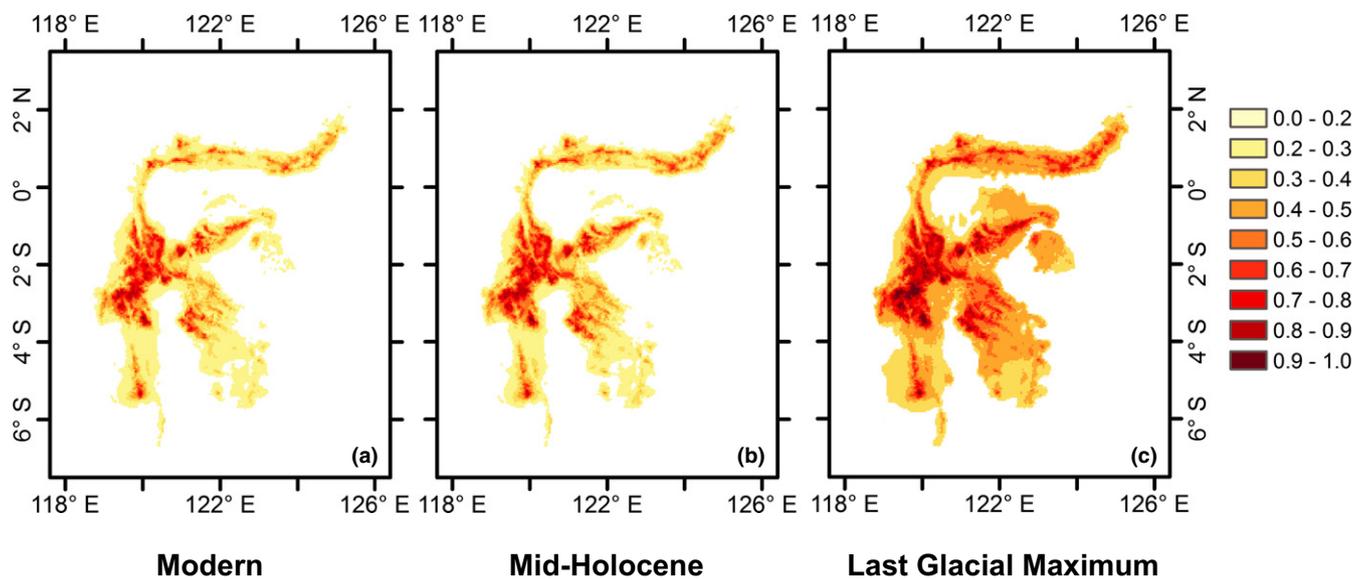


FIGURE 3 Ecological niche models of *Maxomys musschenbroekii* based on eight bioclimatic variables under (a) modern climates, (b) the climate during the mid-Holocene and (c) the climate during the Last Glacial Maximum for Sulawesi, Wallacea. Darker colours indicate areas with higher probability of occurrence

were similar to the modern map (Table S1.6; Schoener's D and Hellinger's $I > 0.953$ in all cases), and there was little difference between circulation models for hindcasted maps (Table S1.7; D and $I > 0.973$ in all cases). Maps of clamping values suggest only the southwestern part of the island was subject to minimal extrapolation, particularly for the CC circulation model at LGM (Fig. S1.8).

3.4 | Isolation by environment and distance

Results of the MMRR analysis comparing environmental distance and over-land geographical distance to genetic distances, while accounting for AoE affiliation, identified a consistent signal of AoE assignment across all eras, regardless of circulation model (Table 4). Regression coefficients (β) of ENM-related isolation were significant for all three eras, but were not significant when using the MPI and MIROC circulation models representing the Last Glacial Maximum (Table 4). Geographical distance coefficients were significant in four of seven models, and the coefficient associated with environmental distance had a larger magnitude than the coefficient associated with geographical distance in 6 of 7 models ($p = 0.125$).

4 | DISCUSSION

4.1 | Phylogeographical structure

Our phylogeographical analyses of *Maxomys musschenbroekii* were based on datasets that differed substantially in data type as well as in the numbers of loci and individuals sampled (Table 2). By analysing diverse data types (e.g., PCNGs vs. UCEs) via multiple methods, we hoped to circumvent potentially confounding data-type effects, which have been shown to bias phylogenetic analyses (Reddy et al., 2017). Our results were largely consistent across analyses of different types of genetic data. Phylogenies produced from CYTB sequences (Figure 1c), PCNGs (Figure 2a), and UCEs (Figure 2b) are similar, with each recovering clades for each of the AoEs and only differing topologically in the placement of the Southeast AoE clade.

One individual from the CYTB tree, MVZ225777, was collected from cluster 13 in the West-Central AoE (Figure 1b) but had a mitochondrial haplotype more closely related to those of individuals from the Southeast AoE (Figure 1c). Given the proximity of cluster 13 to the border between the Southeast and West-Central AoEs, it might be the case that *M. musschenbroekii* has a slightly different position for the border between the West-Central and Southeast AoEs relative to other taxa. This may have been driven by a recent dispersal event. Unlike Eldridge et al. (2018), we did not observe population structuring in relation to elevation, despite having samples spanning at least 1,000 m on nine mountains.

The population structure within *M. musschenbroekii* first uncovered by Achmadi et al. (2013), and further illuminated by our denser sampling of CYTB sequences (Figure 1c), is largely supported by BPP analysis of the PCNGs and UCEs (Table 3; Fig. S1.2). However, there are some disagreements. The West-Central and Southeast AoE clades are delimited consistently across prior schemes and replicates of the PCNG dataset ($PP \geq 0.95$ on average), but not across the prior schemes for the UCE datasets. The Northeast and Southwest AoE clades are delimited across the UCE data subsets and prior schemes, but only by two of the prior schemes used to analyse the PCNG dataset. Two of the AoE clades (East-Central and Northwest) are delimited with high average posterior probabilities by both datasets. Our BPP analyses of UCEs resulted in ranges of posterior probability values among the 40 different combinations of 500-locus subsamples and prior schemes, leading in some cases to conflicting conclusions about population structure. For example, nine of the 40 BPP analyses delimited either or both of the West-Central and Southeast clades as "species" with $PP \geq 0.95$, whereas six did not delimit these clades as "species" ($PP < 0.2$; Table 3; Table S1.4; Fig. S1.2). We only detected this inconsistency in results by sequencing hundreds of loci, replicating analyses with subsamples of the data and testing multiple combinations of priors. Inconsistencies among the BPP results for different prior schemes and datasets likely reflect the relatively low per-locus variation in our nuclear loci (both PCNGs and UCEs; Table 2).

TABLE 4 Results of the multiple matrix regression with randomization (MMRR) analysis assessing correlations between a patristic genetic distance matrix and three other matrices (geographical distances, environmental distances and AoE assignment) for *Maxomys musschenbroekii* in Sulawesi

Era ^a	Scenario	Geographical distance (G)			Environmental distance (E)			E:G β ratio	AoE assignment			Overall model		
		β	t	P-value	β	t	P-value		β	t	P-value	R ²	F	P-value
Modern		-0.015	-3.525	0.064	0.023	5.513	0.004	1.533 ^b	-0.069	-115.115	0.001	0.880	9322.753	0.001
Mid-Hol.	CCSM4	-0.015	-3.743	0.042	0.023	5.807	0.003	1.533 ^b	-0.069	-115.181	0.001	0.880	9331.892	0.001
Mid-Hol.	MPI	-0.034	-8.198	0.001	0.042	10.188	0.001	1.235 ^b	-0.069	-115.928	0.001	0.882	9524.563	0.001
Mid-Hol.	MIROC	-0.019	-4.981	0.010	0.028	7.099	0.001	1.474 ^b	-0.069	-115.505	0.001	0.880	9377.733	0.001
LGM	CCSM4	-0.015	-6.310	0.003	0.023	9.775	0.001	1.533 ^b	-0.068	-113.992	0.001	0.880	9378.225	0.001
LGM	MPI	-0.001	-0.256	0.901	0.009	3.636	0.066	9.000 ^b	-0.069	-113.981	0.001	0.878	9154.467	0.001
LGM	MIROC	0.007	2.822	0.154	0.001	0.505	0.808	0.143	-0.069	-113.481	0.001	0.877	9119.228	0.001

Note. Bold indicates significant value ($P \leq 0.05$).

^aDistances vary under different eras (Modern: 1960–1990; Mid-Holocene: ~6,000 years ago; LGM: ~22,000 years ago), in part, due to shifting sea-levels.

^b β values with higher magnitudes for environmental distance than geographical distance.

Despite some disagreement among datasets and models, our results provide the first evidence that *Maxomys musschenbroekii* could comprise at least six geographically restricted and genetically divergent populations. Because BPP is known to routinely conflate intraspecific population structure with boundaries between species (Sukumaran & Knowles, 2017), we are treating “species delimitations” made by BPP as evidence for population structure under the multispecies coalescent model. Given the relatively shallow divergences among the six clades (all less than about two million years old, Figure 2a), it is unsurprising that AoE clades are not morphometrically distinguishable (Fig. S1.3 and S1.4).

4.2 | Mechanisms of AoE concordance: oceanic and landscape barriers

The phylogeographical structure we found is consistent with the AoEs identified for other species on the island (Evans, Supriatna, Andayani, Setiadi et al., 2003; Fooden, 1969). Given Sulawesi's history as an archipelago comprising up to six precursor islands, marine barriers are an especially plausible mechanism for allopatric divergence in Sulawesi's contemporary endemic terrestrial taxa, with the probability of divergence affected by the depth, width and persistence of the barrier. Once the individual precursor islands became connected, landscape barriers within the growing island may have taken on greater importance.

Because the closest relatives to *M. musschenbroekii* currently inhabit islands on the Sunda shelf (Achmadi et al., 2013), the ancestor to *M. musschenbroekii* likely arrived in proto-Sulawesi via transoceanic sweepstakes dispersal. This process was inferred for up to 80% of the Sulawesi endemic taxa studied by Stelbrink et al. (2012) and is thought to be common for a diversity of terrestrial species over geological time-scales (Heaney, 2007). The alternative, that *M. musschenbroekii* arrived in contemporary Sulawesi via tectonic drift (i.e., dispersing via moving tectonic elements; Michaux, 2010), is unlikely. Rowe, Achmadi, and Esselstyn (2016) used four PCNGs and CYTB to build a fossil-calibrated analysis of SE Asian murines, including 12 *Maxomys* species. They identified the split between *M. musschenbroekii* and its closest relatives as occurring 2 Ma, long after any of the proto-Sulawesi islands could have been connected to Sunda shelf elements like Borneo (Spakman & Hall, 2010; Stelbrink et al., 2012). Depending on the timing of its dispersal across Wallace's Line, the ancestor to *M. musschenbroekii* would have occupied one or more of Sulawesi's precursor islands, all of which were separated by shallow seas (Nugraha & Hall, 2018). If some individuals dispersed to the other precursor islands, this may have initiated the formation of AoE patterns of genetic diversity. Coalescence of the complete modern landmass occurred through uplift between 1 and 4 Ma (Nugraha & Hall, 2018).

We inferred the six AoE clades as having diverged between approximately 230 Ka and 1.5 Ma (Figure 2a), placing the origin of the population structure within *M. musschenbroekii* during the last stages of coalescence and a period of extensive uplift (Nugraha & Hall, 2018). By 1 Ma, only the southern peninsula remained

disconnected from the rest of the island (Nugraha & Hall, 2018). As such, we cannot rule out dispersal across marine barriers as being the mechanism behind the earliest divergences seen within *M. musschenbroekii*. The first inferred divergence split the species into northern (Northeast + Northwest AoE clades) and southern units (the remaining AoE clades) and occurred c. 1.46 Ma (Figure 2a). The second divergence isolated the Southwest clade from a clade that includes the West-Central, East-Central and Southeast AoEs and occurred c. 1.17 Ma (Figure 2a). It is plausible that both of those events were driven by dispersal events among precursor islands. The first dispersal may have occurred near the present-day “neck” of Sulawesi where the northern peninsula connects to the rest of the island, which remained a marine barrier until 1–2 Ma (Nugraha & Hall, 2018). The second cladogenic dispersal event could have occurred across the Tempe Depression, which separates the Southwest peninsula from the rest of the island and which remained a marine barrier up through 1 Ma. That region was periodically inundated during eras of high sea level through the mid-Holocene (Gremmen, 1990). Divergences between the other AoE clades were likely generated by other mechanisms because they most likely occurred after Sulawesi had fully coalesced. Previous authors have noted that some species/populations have contact zones near the Palu-Koro fault zone and Palu River, which runs through the West-Central AoE (e.g., bats and tarsiers; Campbell et al., 2007; Merker et al., 2009). For *M. musschenbroekii*, we observe no obvious divergence between populations on either side of the fault (e.g., clusters 10 vs. 8, 9, 11; Figure 1), but we do not have genetic samples from sites in the immediate vicinity of the fault zone. Overall, while marine barriers are correlated with some of the AoE boundaries, it is possible that other factors have maintained the barriers after seas receded.

4.3 | Mechanisms of AoE concordance: climatic variation and isolation by distance

Climate variation can drive phylogeographical stratification and, eventually, speciation when taxa exhibit phylogenetic niche conservatism, a common pattern wherein closely related species or populations retain the ecological traits (e.g., physiological tolerances) of their common ancestors (Wiens & Graham, 2005; Wiens et al., 2010). For tropical mammals, niche conservatism appears to be strong, especially with regard to temperature tolerances (Cooper, Freckleton, & Jetz, 2011; Olalla-Tárraga et al., 2011). If populations of *M. musschenbroekii* have conserved ecological niches like other tropical mammals, environmental barriers must be considered as potential long-term isolating mechanisms. Isolation due to environmental barriers has, historically, been difficult to distinguish from simple isolation by distance because the two variables are typically correlated. As such, isolation by environment has been neglected as a potential explanatory variable in determining phylogeographical patterns (Wang, 2013). Advances in statistical tools and ecological niche modelling, however, have allowed us to quantify the relative contributions of isolation by environmental suitability and geography on population genetic substructure. On their own, ecological niche



models for *M. musschenbroekii* (Figure 3) illustrate how the geographical pattern of suitable climate space varies across the island and over time, and these areas currently are mostly restricted to mountainous, forested regions. Climate space unsuitable for *M. musschenbroekii* occurs near the boundaries of all AoEs, and sometimes that unsuitable space is extensive (e.g., between the Southwest and West-Central AoEs). Projections of ENMs onto climate models representing the mid-Holocene (~6,000 yr BP) and the LGM (~22,000 yr BP) suggest geographical stability in climate suitability over the last ~20,000 years, regardless of circulation model (Tables S1.7 and S1.8). Because these two eras represent the probable extremes of climates on Sulawesi after coalescence—from cool, dry, low-sea-level periods like the LGM to warm, wet, high-sea-level periods like the mid-Holocene—we can infer that the ecological barriers near AoE boundaries in all three ENMs persisted after the island coalesced in the Pleistocene.

We quantified the effects of isolation by environment and distance using multiple matrix regressions with randomization (MMRR), an analytical tool that can disentangle the effects of multiple variables on genetic distances. In our case, we assessed the extent to which (1) over-land geographical distances, (2) ecological distances inferred from ecological niche model resistance surfaces and (3) AoE assignment could explain patristic distances on the CYTB gene tree. Regardless of era or climate model, our MMRR models (Table 4) have relatively high coefficients of determination (R^2), with the three predictor variables explaining ~88% of the variation in genetic distances across all eras and climate scenarios. The magnitude of the regression coefficients (β) is highest for the AoE assignment matrix, suggesting that the AoE regions defined by toads and macaques (Evans, Supriatna, Andayani, Setiadi et al., 2003) have the strongest predictive value for genetic distances. This result is unsurprising given the obvious geographical structuring apparent in our CYTB phylogeny and is similar to the conclusions of Evans et al. (2004), who used partial Mantel tests to reject isolation by distance in an endemic toad species by demonstrating that individuals within the same AoE were more closely related to those from different AoEs while controlling for geographical distance. In our MMRR analysis, while accounting for AoE assignments, the magnitude of the β values for the environmental distance matrix are consistently larger than those for the geographical distance matrix, suggesting that environmental differences (regardless of era or climate model) better explain differences in genetic distances between AoEs than geographical distance alone. We could not reject that the frequency of larger β values for environmental distances was the result of chance alone, yet this binomial test had low sample size that may preclude a clear result. Overall, these results suggest that environmental differences around AoE boundaries supplemented marine barriers in establishing or maintaining the genetic partitioning we observed.

One potential limitation of this work is the possibility that recent habitat loss has biased the apparent geographical and elevational distribution of *M. musschenbroekii*. Progressive habitat loss that begins in the lowlands and continues uphill, combined with the use of only recently sampled localities for ecological niche modelling, might lead

to the erroneous inference of climatic barriers in lowland areas. We obtained georeferenced collection sites from the online museum collection aggregator VertNet (vertnet.org). This resulted in a compilation of samples collected after 1970 (Table S1.1). The elevational distribution of these samples spans 122–2,300 m elevation. In general, many specimens collected on Sulawesi prior to 1970 (and especially colonial-era samples) are not accessible online, and their descriptive localities are not precise, preventing accurate georeferencing. We suspect that if we could accurately georeference the earliest collections, we would increase the relative representation of lowland sites used for niche models. However, we note that seven of the georeferenced sites used here are from <400 m elevation and these are dispersed widely across the island, being in the general vicinities of site clusters 1, 9 and 10 (Figure 1b). Increases in sample size would affect model fit (Wisz et al., 2008) and expanded geographical representation would affect what would be considered background data (Phillips & Dudík, 2008; Phillips et al., 2006), yielding alternative estimates of climate suitability and its distribution (Anderson & Raza, 2010; Barve et al., 2011). The geographical change from our current models due to such additional data would be difficult to assess *a priori*, but we suspect it would reduce our environmental isolation metrics to values more consistent with geographical isolation.

4.4 | Conclusions

Previous phylogeographical studies of animal taxa on Sulawesi have attributed concordant AoE patterns to landscape features (e.g., marine barriers between palaeo-islands), Pleistocene marine inundations, isolation by distance, or a combination of the three (Driller et al., 2015; Eldridge et al., 2018; Evans, Brown et al., 2003; Evans, Supriatna, Andayani, & Melnick, 2003; Evans, McGuire et al., 2008; Evans, Morales, et al. 2008; Merker et al., 2009; Setiadi et al., 2011; Shelle, Meier, Wahyu, Wirdateti, & Ting, 2010). However, the potential influence of climatic variation as a sole cause or supplement to other causes has not been considered. Overall, our results demonstrate that population structure within *Maxomys musschenbroekii* is remarkably concordant with previously described patterns first noted in macaques and toads and later expanded to include other taxa. For *M. musschenbroekii*, one plausible phylogeographical scenario is that dispersal across marine barriers initiated divergence for the first two intraspecific splits, but were not important for the more recent divergences. For those, climate, more so than isolation by distance, appears to have contributed to isolation between populations. Our results demonstrate that climatic variation near AoE boundaries provides predictive value for genetic divergence, indicating that inhospitable climates may have amplified the effects of short-lived barriers (e.g., marine incursions). This is the first evidence depicting the potential strength of climatic barriers in driving patterns of population structure in a Sulawesi endemic. Similar studies in other Sulawesi endemics could further expose the importance of climate in generating the concordant patterns of genetic partitioning at the centre of the AoE paradigm.

ACKNOWLEDGEMENTS

For assistance with fieldwork, we thank Heru Handika, Jim McGuire, Jim Patton, Carol Patton and Mardin Sarkam. The Indonesian Ministry of Research and Technology, Indonesian Ministry of Environment and Forestry, and Lore Lindu National Park provided permits and logistical support. For access to museum specimens, we thank the Field Museum of Natural History, the Museum of Southwestern Biology and the University of California at Davis Museum of Wildlife and Fish Biology. Portions of this research were conducted with high-performance computing resources provided by Louisiana State University (<http://www.hpc.lsu.edu>). NIH (5R25GM102765-02) funded MKM as she worked on this project. This research was funded by the National Science Foundation (DEB-1441634 and 1343517; OISE-096586), National Geographic Society (9025-11), James and Carol Patton, and a RISTEK travel award to ASA.

ORCID

Thomas C. Giarla  <http://orcid.org/0000-0001-7162-7925>

REFERENCES

- Achmadi, A. S., Esselstyn, J. A., Rowe, K. C., Maryanto, I., & Abdullah, M. T. (2013). Phylogeny, diversity, and biogeography of Southeast Asian spiny rats (*Maxomys*). *Journal of Mammalogy*, *94*, 1412–1423. <https://doi.org/10.1644/13-MAMM-A-092.1>
- Aiello-Lammens, M. E., Boria, R. A., Radosavljevic, A., Vilela, B., & Anderson, R. P. (2015). spThin: an R package for spatial thinning of species occurrence records for use in ecological niche models. *Ecography*, *38*, 541–545. <https://doi.org/10.1111/ecog.01132>
- Anderson, R. P., & Raza, A. (2010). The effect of the extent of the study region on GIS models of species geographic distributions and estimates of niche evolution: Preliminary tests with montane rodents (genus *Nephelomys*) in Venezuela. *Journal of Biogeography*, *37*, 1378–1393. <https://doi.org/10.1111/j.1365-2699.2010.02290.x>
- Arbogast, B. S., Edwards, S. V., Wakeley, J., Beerli, P., & Slowinski, J. B. (2002). Estimating divergence times from molecular data on phylogenetic and population genetic timescales. *Annual Review of Ecology and Systematics*, *33*, 707–740. <https://doi.org/10.1146/annurev.ecolsys.33.010802.150500>
- Arbogast, B. S., & Kenagy, G. J. (2001). Comparative phylogeography as an integrative approach to historical biogeography. *Journal of Biogeography*, *28*, 819–825.
- Barve, N., Barve, V., Jiménez-Valverde, A., Lira-Noriega, A., Maher, S. P., Peterson, A. T., ... Villalobos, F. (2011). The crucial role of the accessible area in ecological niche modeling and species distribution modeling. *Ecological Modelling*, *222*, 1810–1819. <https://doi.org/10.1016/j.ecolmodel.2011.02.011>
- Bellier, O., Sébrier, M., Seward, D., Beaudouin, T., Villeneuve, M., & Putranto, E. (2006). Fission track and fault kinematics analyses for new insight into the Late Cenozoic tectonic regime changes in West-Central Sulawesi (Indonesia). *Tectonophysics*, *413*, 201–220. <https://doi.org/10.1016/j.tecto.2005.10.036>
- Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*, 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
- Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., ... Drummond, A. J. (2014). BEAST 2: A software platform for Bayesian evolutionary analysis. *PLoS Computational Biology*, *10*, e1003537. <https://doi.org/10.1371/journal.pcbi.1003537>
- Bridle, J. R., Pedro, P. M., & Butlin, R. K. (2004). Habitat fragmentation and biodiversity: Testing for the evolutionary effects of refugia. *Evolution*, *58*, 1394. <https://doi.org/10.1111/j.0014-3820.2004.tb01718.x>
- Bynum, E. L., Bynum, D. Z., & Supriatna, J. (1997). Confirmation and location of the hybrid zone between wild populations of *Macaca tonkeana* and *Macaca hecki* in Central Sulawesi, Indonesia. *American Journal of Primatology*, *43*, 181–209. [https://doi.org/10.1002/\(ISSN\)1098-2345](https://doi.org/10.1002/(ISSN)1098-2345)
- Campbell, P., Putnam, A. S., Bonney, C., Bilgin, R., Morales, J. C., Kunz, T. H., & Ruedas, L. A. (2007). Contrasting patterns of genetic differentiation between endemic and widespread species of fruit bats (Chiroptera: Pteropodidae) in Sulawesi, Indonesia. *Molecular Phylogenetics and Evolution*, *44*, 474–482. <https://doi.org/10.1016/j.ympev.2007.02.013>
- Ciani, A. C., Stanyon, R., Scheffrahn, W., & Sampurno, B. (1989). Evidence of gene flow between Sulawesi macaques. *American Journal of Primatology*, *17*, 257–270. [https://doi.org/10.1002/\(ISSN\)1098-2345](https://doi.org/10.1002/(ISSN)1098-2345)
- Cooper, N., Freckleton, R. P., & Jetz, W. (2011). Phylogenetic conservatism of environmental niches in mammals. *Proceedings of the Royal Society B: Biological Sciences*, *278*, 2384–2391. <https://doi.org/10.1098/rspb.2010.2207>
- Darriba, D., Taboada, G. L., Doallo, R., & Posada, D. (2012). jModelTest 2: More models, new heuristics and parallel computing. *Nature Methods*, *9*, 772. <https://doi.org/10.1038/nmeth.2109>
- Demos, T. C., Achmadi, A. S., Giarla, T. C., Handika, H., Rowe, K. C., & Esselstyn, J. A. (2016). Local endemism and within-island diversification of shrews illustrate the importance of speciation in building Sundaland mammal diversity. *Molecular Ecology*, *25*, 5158–5173. <https://doi.org/10.1111/mec.13820>
- 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. <https://doi.org/10.1371/journal.pone.0141212>
- Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*, 1792–1797. <https://doi.org/10.1093/nar/gkh340>
- Eldridge, R. A., Achmadi, A. S., Giarla, T. C., Rowe, K. C., & Esselstyn, J. A. (2018). Geographic isolation and elevational gradients promote diversification in an endemic shrew on Sulawesi. *Molecular Phylogenetics and Evolution*, *118*, 306–317. <https://doi.org/10.1016/j.ympev.2017.09.018>
- Esselstyn, J. A., Maharadatunkamsi, Achmadi, A. S., Siler, C. D., & Evans, B. J. (2013). Carving out turf in a biodiversity hotspot: Multiple, previously unrecognized shrew species co-occur on Java Island, Indonesia. *Molecular Ecology*, *22*, 4972–4987. <https://doi.org/10.1111/mec.12450>
- ESRI (2012). *ArcGIS Desktop: Release 10*. Redlands, CA: Environmental Systems Research Institute.
- Evans, B. J., Brown, R. M., McGuire, J. A., Supriatna, J., Andayani, N., Diesmos, A., ... Cannatella, D. C. (2003). Phylogenetics of fanged frogs: Testing biogeographical hypotheses at the interface of the Asian and Australian faunal zones. *Systematic Biology*, *52*, 794–819.
- Evans, B. J., Cannatella, D. C., & Melnick, D. J. (2004). Understanding the origins of areas of endemism in phylogeographic analyses: A reply to Bridle et al. *Evolution*, *58*, 1397–1400.
- Evans, B. J., McGuire, J. A., Brown, R. M., Andayani, N., & Supriatna, J. (2008). A coalescent framework for comparing alternative models of population structure with genetic data: Evolution of Celebes toads. *Biology Letters*, *4*, 430–433. <https://doi.org/10.1098/rsbl.2008.0166>
- Evans, B. J., Morales, J. C., Supriatna, J., & Melnick, D. J. (2008). Origin of the Sulawesi macaques (Cercopithecidae: *Macaca*) as suggested by



- mitochondrial DNA phylogeny. *Biological Journal of the Linnean Society*, 66, 539–560.
- Evans, B. J., Supriatna, J., Andayani, N., & Melnick, D. J. (2003). Diversification of Sulawesi macaque monkeys: Decoupled evolution of mitochondrial and autosomal DNA. *Evolution*, 57, 1931–1946. <https://doi.org/10.1111/j.0014-3820.2003.tb00599.x>
- Evans, B. J., Supriatna, J., Andayani, N., Setiadi, M. I., Cannatella, D. C., & Melnick, D. J. (2003). Monkeys and toads define areas of endemism on Sulawesi. *Evolution*, 57, 1436–1443. <https://doi.org/10.1111/j.0014-3820.2003.tb00350.x>
- Faircloth, B. C. (2016). PHYLUCS is a software package for the analysis of conserved genomic loci. *Bioinformatics*, 32, 786–788. <https://doi.org/10.1093/bioinformatics/btv646>
- Faircloth, B. C., McCormack, J. E., Crawford, N. G., Harvey, M. G., Brumfield, R. T., & Glenn, T. C. (2012). Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary time-scales. *Systematic Biology*, 61, 717–726. <https://doi.org/10.1093/sysbio/sys004>
- Fooden, J. (1969). Taxonomy and evolution of the monkeys of Celebes (Primates: Cercopithecidae). *Bibliotheca Primatologica*, 10, 1–148.
- Fourment, M., & Gibbs, M. J. (2006). PATRISTIC: A program for calculating patristic distances and graphically comparing the components of genetic change. *BMC Evolutionary Biology*, 6, 1. <https://doi.org/10.1186/1471-2148-6-1>
- Giarla, T. C., & Esselstyn, J. A. (2015). The challenges of resolving a rapid, recent radiation: Empirical and simulated phylogenomics of Philippine shrews. *Systematic Biology*, 64, 727–740. <https://doi.org/10.1093/sysbio/syv029>
- Gillespie, R. (2004). Community assembly through adaptive radiation in Hawaiian spiders. *Science*, 303, 356–359. <https://doi.org/10.1126/science.1091875>
- Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., ... Zeng, Q. (2011). Trinity: Reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nature Biotechnology*, 29, 644–652. <https://doi.org/10.1038/nbt.1883>
- Grant, B. R., & Grant, P. R. (2008). Fission and fusion of Darwin's finches populations. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 363, 2821–2829. <https://doi.org/10.1098/rstb.2008.0051>
- Gremmen, W. (1990). Palynological investigations in the Danau Tempe depression, southwest Sulawesi (Celebes), Indonesia. *Modern Quaternary Research in Southeast Asia*, 11, 123–134.
- Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Systematic Biology*, 59, 307–321. <https://doi.org/10.1093/sysbio/syq010>
- Hall, R. (2002). Cenozoic geological and plate tectonic evolution of SE Asia and the SW Pacific: Computer-based reconstructions, model and animations. *Journal of Asian Earth Sciences*, 20, 353–431. [https://doi.org/10.1016/S1367-9120\(01\)00069-4](https://doi.org/10.1016/S1367-9120(01)00069-4)
- Hardy, D. K., González-Cózar, F. X., Arellano, E., & Rogers, D. S. (2013). Molecular phylogenetics and phylogeographic structure of Sumichrast's harvest mouse (*Reithrodontomys sumichrasti*: Cricetidae) based on mitochondrial and nuclear DNA sequences. *Molecular Phylogenetics and Evolution*, 68, 282–292. <https://doi.org/10.1016/j.ympev.2013.03.028>
- Harvey, M. G., Smith, B. T., Glenn, T. C., Faircloth, B. C., & Brumfield, R. T. (2016). Sequence capture versus restriction site associated DNA sequencing for shallow systematics. *Systematic Biology*, 65, 910–924. <https://doi.org/10.1093/sysbio/syw036>
- Heaney, L. R. (2007). Is a new paradigm emerging for oceanic island biogeography? *Journal of Biogeography*, 34, 753–757. <https://doi.org/10.1111/j.1365-2699.2007.01692.x>
- Heaney, L. R., Balete, D. S., Duya, M. R. M., Duya, M. V., Jansa, S. A., Steppan, S. J., & Rickart, E. A. (2016). Doubling diversity: A cautionary tale of previously unsuspected mammalian diversity on a tropical oceanic island. *Frontiers of Biogeography*, 8, 1–19.
- Heaney, L. R., Balete, D. S., Rickart, E. A., Alviola, P. A., Duya, M. R. M., Duya, M. V., & Steppan, S. J. (2011). Chapter 1: Seven new species and a new subgenus of forest mice (Rodentia: Muridae: *Apomys*) from Luzon island. *Fieldiana Life and Earth Sciences*, 2, 1–60. <https://doi.org/10.3158/2158-5520-2.1.1>
- Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, 25, 1965–1978. [https://doi.org/10.1002/\(ISSN\)1097-0088](https://doi.org/10.1002/(ISSN)1097-0088)
- Hijmans, R. J., & vanEtten, J. (2014). *raster: Geographic data analysis and modeling. R package version 2.6-7*.
- Hijmans, R. J., Phillips, S., Leathwick, J., & Elith, J. (2015). *dismo: Species distribution modeling. R package version 1.1-4*.
- Irwin, D. E. (2002). Phylogeographic breaks without geographic barriers to gene flow. *Evolution*, 56, 2383–2394. <https://doi.org/10.1111/j.0014-3820.2002.tb00164.x>
- Justiniano, R., Schenk, J. J., Balete, D. S., Rickart, E. A., Esselstyn, J. A., Heaney, L. R., & Steppan, S. J. (2014). Testing diversification models of endemic Philippine forest mice (*Apomys*) with nuclear phylogenies across elevational gradients reveals repeated colonization of isolated mountain ranges. *Journal of Biogeography*, 42, 51–64.
- Lack, D. (1947). *Darwin's Finches*. Cambridge: Cambridge University Press.
- Lohman, D. J., de Bruyn, M., Page, T., von Rintelen, K., Hall, R., Ng, P. K. L., ... von Rintelen, T. (2011). Biogeography of the Indo-Australian archipelago. *Annual Review of Ecology, Evolution, and Systematics*, 42, 205–226. <https://doi.org/10.1146/annurev-ecolsys-102710-145001>
- Losos, J., Jackman, T., Larson, A., Queiroz, K., & Rodriguez-Schettino, L. (1998). Contingency and determinism in replicated adaptive radiations of island lizards. *Science*, 279, 2115–2118. <https://doi.org/10.1126/science.279.5359.2115>
- Merker, S., Driller, C., Perwitasari-Farajallah, D., Pamungkas, J., & Zischler, H. (2009). Elucidating geological and biological processes underlying the diversification of Sulawesi tarsiers. *Proceedings of the National Academy of Sciences*, 106, 8459–8464. <https://doi.org/10.1073/pnas.0900319106>
- Michaux, B. (2010). Biogeology of Wallacea: Geotectonic models, areas of endemism, and natural biogeographical units. *Biological Journal of the Linnean Society*, 101, 193–212. <https://doi.org/10.1111/j.1095-8312.2010.01473.x>
- Miller, M. A., Pfeiffer, W., & Schwartz, T. (2010). Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *Gateway Computing Environments Workshop (GCE)*, 2010, 1–8.
- Muscarella, R., Galante, P. J., Soley-Guardia, M., Boria, R. A., Kass, J. M., Uriarte, M., & Anderson, R. P. (2014). ENMeval: An R package for conducting spatially independent evaluations and estimating optimal model complexity for Maxent ecological niche models. *Methods in Ecology and Evolution*, 5, 1198–1205. <https://doi.org/10.1111/2041-210X.12261>
- Nugraha, A. M. S., & Hall, R. (2018). Late Cenozoic palaeogeography of Sulawesi, Indonesia. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 490, 191–209. <https://doi.org/10.1016/j.palaeo.2017.10.033>
- Ogilvie, H. A., Bouckaert, R. R., & Drummond, A. J. (2017). StarBEAST2 brings faster species tree inference and accurate estimates of substitution rates. *Molecular Biology and Evolution*, 34, 2101–2114. <https://doi.org/10.1093/molbev/msx126>
- Olalla-Tárraga, M. Á., McInnes, L., Bini, L. M., Diniz-Filho, J. A. F., Fritz, S. A., Hawkins, B. A., ... Purvis, A. (2011). Climatic niche conservatism and the evolutionary dynamics in species range boundaries: Global congruence across mammals and amphibians. *Journal of Biogeography*, 38, 2237–2247. <https://doi.org/10.1111/j.1365-2699.2011.02570.x>
- Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 190, 231–259. <https://doi.org/10.1016/j.ecolmodel.2005.03.026>

- Phillips, S. J., & Dudík, M. (2008). Modeling of species distributions with Maxent: New extensions and a comprehensive evaluation. *Ecography*, *31*, 161–175. <https://doi.org/10.1111/j.0906-7590.2008.5203.x>
- R Core Team. (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>
- Reddy, S., Kimball, R. T., Pandey, A., Hosner, P. A., Braun, M. J., Hackett, S. J., ... Braun, E. L. (2017). Why do phylogenomic data sets yield conflicting trees? Data type influences the avian tree of life more than taxon sampling. *Systematic Biology*, *66*, 857–879. <https://doi.org/10.1093/sysbio/syx041>
- Rowe, K. C., Achmadi, A. S., & Esselstyn, J. A. (2016). Repeated evolution of carnivory among Indo-Australian rodents. *Evolution*, *70*, 653–665. <https://doi.org/10.1111/evo.12871>
- Setiadi, M. I., McGuire, J. A., Brown, R. M., Zubairi, M., Iskandar, D. T., Andayani, N., ... Evans, B. J. (2011). Adaptive radiation and ecological opportunity in Sulawesi and Philippine fanged frog (*Limnonectes*) communities. *The American Naturalist*, *178*, 221–240. <https://doi.org/10.1086/660830>
- Shah, V. B., & McRae, B. H. (2008). Circuitscape: A tool for landscape ecology. *Proceedings of the 7th Python in Science Conference*, *7*, 62–66.
- Shekelle, M., Meier, R., Wahyu, I., Wirdateti, & Ting, N. (2010). Molecular phylogenetics and chronometrics of Tarsiidae based on 12s mtDNA haplotypes: Evidence for Miocene origins of crown tarsiers and numerous species within the Sulawesi clade. *International Journal of Primatology*, *31*, 1083–1106. <https://doi.org/10.1007/s10764-010-9457-8>
- Spakman, W., & Hall, R. (2010). Surface deformation and slab–mantle interaction during Banda arc subduction rollback. *Nature Geoscience*, *3*, 562–566. <https://doi.org/10.1038/ngeo917>
- Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*, 1312–1313. <https://doi.org/10.1093/bioinformatics/btu033>
- Stelbrink, B., Albrecht, C., Hall, R., & von Rintelen, T. (2012). The biogeography of Sulawesi revisited: Is there evidence for a vicariant origin of taxa on Wallace's 'anomalous island'? *Evolution*, *66*, 2252–2271. <https://doi.org/10.1111/j.1558-5646.2012.01588.x>
- Stephens, M., Smith, N. J., & Donnelly, P. (2001). A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics*, *68*, 978–989. <https://doi.org/10.1086/319501>
- Sukumaran, J., & Knowles, L. L. (2017). Multispecies coalescent delimits structure, not species. *Proceedings of the National Academy of Sciences*, *114*, 1607–1612. <https://doi.org/10.1073/pnas.1607921114>
- Toussaint, E. F. A., Hall, R., Monaghan, M. T., Sagata, K., Ibalim, S., Shaverdo, H. V., ... Balke, M. (2014). The towering orogeny of New Guinea as a trigger for arthropod megadiversity. *Nature Communications*, *5*, 4001. <https://doi.org/10.1038/ncomms5001>
- Wang, I. J. (2013). Examining the full effects of landscape heterogeneity on spatial genetic variation: A multiple matrix regression approach for quantifying geographic and ecological isolation. *Evolution*, *67*, 3403–3411. <https://doi.org/10.1111/evo.12134>
- Warren, D. L., & Seifert, S. N. (2011). Ecological niche modeling in Maxent: The importance of model complexity and the performance of model selection criteria. *Ecological Applications*, *21*, 335–342. <https://doi.org/10.1890/10-1171.1>
- Wiens, J. J., Ackerly, D. D., Allen, A. P., Anacker, B. L., Buckley, L. B., Cornell, H. V., ... Stephens, P. R. (2010). Niche conservatism as an emerging principle in ecology and conservation biology. *Ecology Letters*, *13*, 1310–1324. <https://doi.org/10.1111/j.1461-0248.2010.01515.x>
- Wiens, J. J., & Graham, C. H. (2005). Niche conservatism: Integrating evolution, ecology, and conservation biology. *Annual Review of Ecology, Evolution, and Systematics*, *36*, 519–539. <https://doi.org/10.1146/annurev.ecolsys.36.102803.095431>
- Wis, M.S., Hijmans, R. J., Li, J., Peterson, A. T., Graham, C. H., & Guisan, A.; NCEAS Predicting Species Distributions Working Group. (2008). Effects of sample size on the performance of species distribution models. *Diversity and Distributions*, *14*, 763–773. <https://doi.org/10.1111/j.1472-4642.2008.00482.x>
- Wu, C. I., & Li, W. H. (1985). Evidence for higher rates of nucleotide substitution in rodents than in man. *Proceedings of the National Academy of Sciences*, *82*, 1741–1745. <https://doi.org/10.1073/pnas.82.6.1741>
- Yang, Z. (2015). The BPP program for species tree estimation and species delimitation. *Current Zoology*, *61*, 854–865. <https://doi.org/10.1093/czoolo/61.5.854>
- Yang, Z., & Rannala, B. (2010). Bayesian species delimitation using multi-locus sequence data. *Proceedings of the National Academy of Sciences*, *107*, 9264–9269. <https://doi.org/10.1073/pnas.0913022107>
- Yang, Z., & Rannala, B. (2014). Unguided species delimitation using DNA sequence data from multiple Loci. *Molecular Biology and Evolution*, *31*, 3125–3135. <https://doi.org/10.1093/molbev/msu279>

DATA ACCESSIBILITY

Protein-coding DNA sequences are available on GenBank under accession codes MH068879–MH069205. Raw ultraconserved element (UCE) reads and DNA sequences are available in the NCBI BioProject PRJNA437989.

BIOSKETCH

Thomas Giarla is an Assistant Professor of Biology at Siena College. He is interested in the geographical and climatological forces that drive diversification and primarily studies small mammals from the tropics. He and the other authors on this paper work to understand the ecology, diversity, and historical biogeography of small mammals from Southeast Asia.

Author contributions: T.C.G. and J.A.E. conceived the ideas for this study. J.A.E., K.C.R., A.S.A. and M.T.S. conducted the fieldwork. T.C.G., M.K.M. and M.T.S. conducted the laboratory work. T.C.G. conducted the genetic analyses, J.A.E. conducted the morphological analyses, and S.P.M. conducted the niche modelling and regression analyses. T.C.G. led the writing of the manuscript, with assistance from J.A.E. and S.P.M.

SUPPORTING INFORMATION

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

How to cite this article: Giarla TC, Maher SP, Achmadi AS, et al. Isolation by marine barriers and climate explain areas of endemism in an island rodent. *J Biogeogr.* 2018;00:1–14. <https://doi.org/10.1111/jbi.13392>