Expanding Population Edge Craniometrics and Genetics Provide Insights into Dispersal of Commensal Rats through Nusa Tenggara, Indonesia

Julien Louys[1] (iD), Michael B. Herrera[2,3], Vicki A. Thomson[2] (iD), Andrew S. Wiewel[2] (iD), Stephen C. Donnellan[2,4] (iD), Sue O'Connor[5,6] (iD), and Ken Aplin[7]† (iD)

[1] Australian Research Centre for Human Evolution, Griffith University, Brisbane QLD 4111, Australia

[2] School of Biological Sciences, University of Adelaide, Adelaide SA 5005, Australia

[3] Archaeological Studies Program, University of the Philippines, Diliman, Quezon City, 1101, Manila, Philippines

[4] South Australian Museum, North Terrace, Adelaide SA 5000, Australia

[5] Archaeology and Natural History, College of Asia and the Pacific,
The Australian National University, Acton ACT 2601, Australia

[6] ARC Centre of Excellence for Australian Biodiversity and Heritage,
The Australian National University, Canberra, ACT 2601, Australia

[7] Australian Museum Research Associate, 1 William Street, Sydney NSW 2010, Australia

Abstract. The Nusa Tenggara island chain consists of an archipelago that runs roughly east-west in eastern Indonesia. As part of Wallacea, it has never been connected to any continental landmass, and has been subject to a variety of biological invasions that have populated the islands. Here, we examine the craniometric and molecular genetic records of several species of *Rattus sensu lato* in the island chain. We use the predictions of expanding population edge phenotypic selection in an effort to understand the movement of *Rattus rattus* and *Rattus exulans* through the archipelago. We also examine the mitochondrial haplotype networks of *R. argentiventer, R. exulans*, and the *R. rattus* Complex (RrC) and microsatellite allele frequency clustering patterns for the RrC*,* to examine relationships within and between Nusa Tenggara populations, and those of Asia and the Pacific where relevant for each taxon. In the RrC LIV and RrC LII haplotype networks, 20 haplotypes with seven from Nusa Tenggara were observed for RrC LIV, and 100 haplotypes with seven from Nusa Tenggara observed for RrC LII. The top performing RrC craniometric model had a negative association between size and distance from the easternmost point of the samples from Nusa Tenggara, consistent with increasing size moving west to east. The *cytochrome b* network for the *R. exulans* sequences comprised 14 haplotypes, with three observed from mainland Southeast Asia, one shared with Nusa Tenggara and regions further east, and another haplotype observed in Nusa Tenggara and in the Pacific. The *R. exulans* craniometric model selection produced four equally well performing models, with no migration scenario preferred. Finally, the haplotype network of *R. argentiventer* comprised 10 haplotypes, with six observed in Nusa Tenggara, including a relatively early cluster from the east of the archipelago. Our results are compatible with a polyphasic and polydirectional invasion of Nusa Tenggara by *Rattus*, likely beginning with RrC from the west to the east, an expansion of *R. exulans* from Flores, seemingly in no preferred overall direction, and finally the invasion of *R. argentiventer* from the east to the west. We find some support for the Dong Son drum maritime exchange network contributing to the distribution of the latter species.

† Ken Aplin 1958–2019, deceased

Introduction

The expansion of species outside of their initial core range is associated with phenotypic and genotypic changes resulting from the dynamics of expansion (Chuang & Peterson, 2016). Differences largely arise between edge and core demes (subpopulations) as a result of isolation, changing environmental gradients, and physiological, metabolic, and behavioural demands of dispersal (Chuang & Peterson, 2016). The most commonly observed dispersal-promoting phenotypic trait is related to larger size or better body condition (Chuang & Peterson, 2016). Thus, a prediction from expanding populations edge studies is that individuals at the invasion front of a species will be bigger and better conditioned than those at the core. The classic example of this phenomenon is the cane toad in Australia, where individuals at the leading edge of the invasion are longer legged than at the core, thereby promoting further and faster dispersal through time (Phillips *et al*., 2006, 2007). Genetically, the most commonly expected result of dispersal into islands are reduced gene flow from core to the invasion front, resulting in the evolution of unique genetic markers in relatively isolated island populations. Arguably, one of the most important biological expansion events, at least from a human perspective (and other than our own), involved the genus *Rattus*.

Rattus is one of the most speciose mammalian genera, and one of the most impactful on human health and subsistence (Aplin *et al*., 2003). Despite this, its biogeographic and evolutionary history remains poorly known. While the general geographic origin of several of the most important commensal taxa have been relatively well established (Aplin *et al*., 2011; Thomson *et al*., 2014), how and in which directions they spread from their core areas, and how they interacted with each other is still poorly resolved. The Neolithic introduction of rodents through Southeast Asia and into the Pacific are probably the first and most well-studied of these commensal invasions. They are thought to be part of an agriculturalist "package" that also included the introduction of chickens and pigs, the widespread adoption of agricultural practices, and involved species such as the Pacific Rat *Rattus exulans* and the Black Rat *R. rattus* (Thomson *et al.*, 2018; Louys *et al*., 2018a; Leppard, 2018). The importation of pest species throughout the Asia-Pacific likely produced significant impacts on human subsistence and populations, particularly through raiding of crop stores, by acting as disease vectors, and destroying island ecosystems (Aplin *et al*., 2003, 2011; Leppard, 2018).

The islands of Nusa Tenggara provide an excellent region to study colonization and diversification of commensal rodents. The Nusa Tenggara archipelago represents a series of oceanic islands running in a roughly east-west orientation in the Wallacean region of island Southeast Asia (ISEA). Emerging probably sometime in the Pliocene, fauna dispersed to these islands in a series of colonization events, the most recent of which was tied to deliberate or unintentional introductions of commensals by humans. Invasions of rodents into the island chain continued into the Metal Age and possibly well into historic times (St Pierre, 2011). These islands are the likely point of origin for the Pacific Rat *Rattus exulans* (Thomson *et al*., 2014) and have been linked with each other through maritime trade networks since at least the Pleistocene (Reepmeyer *et al.*, 2016), and with mainland Southeast Asia (MSEA) since the Neolithic (Bellwood, 2007). Several important commensal species are currently found on the islands; besides *R. exulans* they include the Black Rat *R. rattus* and the Ricefield Rat *R. argentiventer*.

Ecological studies of shifting population ranges have often studied biological invasions (Chuang & Peterson, 2016). The introduction and dispersal of commensal rodents through Nusa Tenggara, even if anthropologically facilitated, might share ecological characteristics with these. We approached the question of movement into and through Nusa Tenggara through two independent but complementary approaches.

Firstly, we examined molecular genetic data of three rodent species belonging to *Rattus sensu lato*. The taxonomic identity of some commensal rodent taxa is beset with issues of unresolved taxonomic resolution in the natural range. In some cases these issues are amplified in the introduced range, due to multiple introductions from diverse source populations, and in some cases further confounded by subsequent introgression in the introduced range. The *Rattus rattus* Complex (RrC hereafter) is one group where all of the above factors are present in several parts of the natural and introduced ranges (Aplin *et al*., 2011; Pages *et al*., 2013; Lack *et al*., 2012). In particular, mitochondrial lineages RrC LII and LIV co-occur in Indonesia and likely have dispersed there from MSEA and the Philippines (Aplin *et al*., 2011). While these lineages freely interbreed in Indochina (Pages *et al*., 2013) and show introgressive hybridization in the Philippines (Lack *et al*., 2012), their status as separate evolutionary entities in Indonesia is not known and could critically affect the interpretation of dispersal history in the archipelago. Thus, part of our study presents novel molecular genetic data informing on the identity of the RrC in Indonesia. Alongside RrC, we also examine molecular genetic data on *R. exulans* and *R. argentiventer*.

Secondly, we examine the craniometrics of three species of *Rattus* to examine what changes, if any, can be observed between populations on different islands. We examine whether there is a trend in increasing size from the core to the population edge of a species in the commensal *R. rattus* and *R. exulans*, which may owe part or all their expansion to anthropogenic influence. Such a trend, if present, might relate to the selection effects of expansion. The two taxa are shared in common with the genetic analyses. We also included one wild taxon, *R. hainaldi* from Flores. *Rattus hainaldi,* although currently classified in *Rattus,* is phylogenetically associated with a group that includes *Tarsomys, Limnomys, Diplothrix, Nesokia,* and *Bandicota,* and revision of *Rattus* may necessitate the removal of *hainaldi* from this genus (Thomson *et al*., 2018). Nevertheless, it is important as an endemic rodent from Flores that, despite being found alongside *R. exulans*, has remained wild and has not dispersed beyond its island (Kitchener *et al*., 1991; Veatch *et al*., 2019). It thus provides an important indicator of cranial metric variation in a restricted island endemic. Finally, we discuss our results with a consideration of the archaeological record of maritime trade and human dispersal between Nusa Tenggara and the broader Asia-Pacific region.

Materials and methods

Mitochondrial DNA sequencing and analysis

Building haplotype networks across an island chain and including extralimital samples can provide important insights into directions of gene flow, as unique haplogroups are traced through successive islands (or geographic regions). Such analyses have been used to inform on the movements of RrC through Southeast Asia (Aplin *et al.*, 2011) and the likely direction of invasion for RrC into the Talaud islands (Louys *et al.*, 2018a). The specimens used in this study included museum tissues and remains collected from surface finds (Table S1, see Louys *et al.*, 2020). Genomic DNA was extracted from museum tissue samples using a salting-out method (Nicholls & Austin, 2005). Surface remains were extracted in specialist ancient DNA (aDNA) laboratories of the Australian Centre for Ancient DNA (ACAD) at the University of Adelaide. DNA was extracted from each rodent incisor or bone element using the DNeasy Kit (Qiagen) with modification following the procedure described in Thomson *et al.* (2014).

A total of 380 samples were examined for cytochrome *b* (*cyt b*) sequences across four taxa: *R. argentiventer* (n = 39), *R. exulans* (n = 20); RrC LIV (n = 49); and RrC LII (n = 272). All other sequences were sourced from the literature (Table S1, see Louys *et al.*, 2020). The PCR primers used to generate the mitochondrial *cyt b* sequences are listed in Table S1, see Louys *et al.*, 2020 (ACAD835, ACAD 1936, ACAD1937, and ACAD1877). PCR reactions were set up in 25 µL volumes containing a final concentration of $1 \times$ HiFi PCR buffer (Platinum Invitrogen), 200 µM each dNTP, 3 mM MgSO4, 1 mg/mL^{-1} Rabbit Serum Albumin (Sigma), 1 µM of each primer, 1 unit of Platinum *Taq*HiFi DNA polymerase (Invitrogen), and 2 µl of template DNA. Thermocycling included initial denaturation and enzyme activation at 94°C for 2 minutes, then 55 cycles of denaturing at 94°C for 30 seconds, primer annealing at 55°C for 30 seconds and extension at 68°C for 30 seconds, and a final extension 68°C for 10 minutes. Amplifications of extractions and PCR blank controls were also performed in all experiments to monitor for contamination. Amplicons were separated by electrophoresis on a 2.5% agarose gel. PCR clean-up, Sanger sequencing and capillary electrophoresis were conducted at the Australian Genome Research Facility Ltd (Australia).

The forward and reverse sequence chromatograms were aligned, visually inspected and edited using Geneious v. 7.1.2. (Biomatters) to obtain a consensus sequence. The newly generated sequences were aligned with published sequences (Table S1, see Louys *et al.*, 2020) using MUSCLE alignment algorithm (Edgar, 2004) to form the *cyt b* dataset. The evolutionary relationships among the haplotypes were characterized using median-joining (MJ) network analysis Popart v. 1.7.1 (Leigh & Bryant, 2015).

Microsatellite genotyping and analysis

Prospective RrC microsatellite loci were identified by shotgun sequencing of two samples (RrC I-ABTC050177; RrC III-ABTC109244) on the GS-FLX platform (454 Life Sciences/Roche FLX) at AGRF-SA, following the protocol in Gardner *et al.* (2011). The resulting sequences were screened for microsatellite repeats using Msatcommander v. 1.0.8 (Faircloth, 2008) and checked for unique flanking sequences

using MICROFAMILY (Meglécz, 2007), resulting in 395 potentially useful microsatellite repeats (194 dinucleotide, 79 trinucleotide, 109 tetranucleotide, 8 pentanucleotide, and 5 hexanucleotide). From this list, 30 microsatellite loci (1 dinucleotide, 7 trinucleotide, 21 tetranucleotide, and 1 hexanucleotide) were selected for primer development.

Primer3 (Rozen & Skaletsky, 2000) was used to design microsatellite primers with Multiplex Ready Technology (MRT) tag sequences to facilitate flexible fluorescent labelling (FAM, NED, PET, and VIC) and post-PCR multiplexing (Hayden *et al.*, 2008). These primers were evaluated to determine the reliability of PCR amplification and presence of polymorphism across multiple RrC lineages, and to determine the optimum primer concentration (10, 20, 40 or 60 nM) for each primer pair, resulting in a final selection of 12 microsatellite loci (Table 1) that were pooled into two multiplexes.

Our 12 microsatellite loci were PCR amplified for 174 natural and introduced range RrC lineage II (n = 79) and IV (n = 95) samples in a volume of 12 µL containing 2.4 µL 5X MRT buffer (1.2 µL Immolase buffer, 0.36 µL MgCl$_2$ (50 mM), 0.096 µL dNTPs (100 mM), 0.06 µL BSA (100X), and 0.684 µL ddH$_2$0), 0.06 µL Immolase (5U/µL), 0.09 µL fluorescent-labelled forward tag (10 µM), 0.09 µL unlabelled reverse tag (10 µM), 0.03–0.06 µL of the locus-specific primer pair (4 µM), 7.3–7.33 µL ddH$_2$O, and 2 µL template DNA (diluted to introduce 10–20 ng DNA) (Table 1). The MRT PCR thermocycling profile consisted of an initial denaturation step of 95°C for 10 min, 5 cycles of denaturation at 92°C for 60 s, annealing at 50°C for 90 s, and extension at 72°C for 60 s, 20 cycles of denaturation at 92°C for 30 s, annealing at 63°C for 90 s, and extension at 72°C for 60 s, 40 cycles of denaturation at 92°C for 15 s, annealing at 54°C for 60 s, and extension at 72°C for 60 s, and a final extension at 72°C for 10 min. Two control samples were included in every 96-well plate to ensure consistent amplification between runs, and eight samples were repeated in reverse order within each plate to identify plate orientation errors and to calculate genotyping error rate.

Fluorescent-labelled PCR products from each multiplex were pooled by DNA sample with a ratio of 3 FAM: 2 NED: 4 PET: 2.25 VIC to account for differences in relative fluorescence of each dye, cleaned using MultiScreen® PCR cleanup filter plates (Millipore Corporation, Billerica, MA, USA), and analyzed using an ABI 3730xl DNA Analyzer (Applied Biosystems) at AGRF-SA. Microsatellite alleles were visualized and manually scored using GeneMapper v. 3.7 (Applied Biosystems), with a subset of genotypes confirmed by an independent party to minimize bias. One microsatellite performed poorly (amplification for c. 50% of samples) and was excluded from further analyses.

Microsatellite loci were tested for deviations from Hardy-Weinberg equilibrium and linkage disequilibrium using GENEPOP v. 4.1 (Rousset, 2008), with Bonferroni correction for multiple comparisons. Diversity statistics, including allelic richness (A), observed and expected heterozygosity (H$_O$ and H$_E$, respectively), and inbreeding coefficient (F$_{IS}$), were calculated in GenAlEx v. 6.5 (Peakall & Smouse, 2012). We investigated broad-scale genetic structure in our RrC lineage II and IV sample by conducting a principal coordinates analysis (PCoA) in GenAlEx v. 6.5 (Peakall & Smouse, 2012). Finally, we used the Bayesian clustering approach implemented in Structure v. 2.3.4

Table 1. Microsatellite locus name, primer sequences, fluorescent label, repeat motif, and amplicon size at primer testing.

| locus	primer sequence (5'–3')	label	repeat motif	size range (bp)
Rrat6	F: CACCAGTGTCCAATAACTATCCG			
	R: TGAATTGCTATAAGTGGGTAAAAGA	FAM	$(AGAT)_{14}$	107
Rrat8	F: AGGGCTTTTGTGGGTTTGTT			
	R: GGACAACAGGAGGCCTTGTC	NED	$(AGAT)_{15}$	111
Rrat11	F: CCAATGCCCAGCAAGATTTA			
	R: ACATGGCTCAAGGCATACAT	VIC	$(ACAT)_{13}$	128
Rrat15	F: TCCTTAGGTGTCAACAGCACTC			
	R: TCTGCACTCTTGACTCCAACA	VIC	$(AGAT)_{13}$	152
Rrat17	F: GAAGCCACATTTACCCCTGA			
	R: GTTGCCTAGTTTGCCTTGGA	PET	$(AAAT)_{14}$	167
Rrat21	F: CCAGCACTTGGGAGGTAAAA			
	R: TGTTCAAACCAGCCTTCTCA	NED	$(AAGC)_{12}$	199
Rrat24	F: GCGCCCTGTGTCTTACTGTT			
	R: CCAGAAGCTAATATAGAAATGTGGC	FAM	$(AAAT)_{11}$	218
Rrat25	F: TGTGCATGGAGTGCCTTCTA			
	R: TTGTCAAGGTTAGTGACTACTTTCC	VIC	$(ACT)_{17}$	234
Rrat26	F: TTTAACAGCGGAGGAGCAGT			
	R: ACTAACTGCAATTCGTGGGG	PET	$(AAGC)_{11}$	234
Rrat27	F: AGTCAGAAGCAAACCAGCGT			
	R: GAAACCAATTCCAAAACACTCA	NED	$(AAAT)_{12}$	261
Rrat28	F: GAGCTGCTGTCTTTCCATCC			
	R: CCCATGAAATCTCAAAGGTATG	PET	$(AAC)_{15}$	281
Rrat30	F: CCCAAGAACTGAGCAAGAGG			
	R: TGAATGGCCTATAACCACAACTT	FAM	$(AGAT)_{9}$	316

(Pritchard *et al.*, 2000) to infer the optimal number of genetic clusters (K) in our dataset and to assign individuals to those clusters. Structure simulations were parameterized to run 20 independent analyses of 1 million generations following a 500,000 generation burn-in for K = 1 to K = 9 (the total number of countries in our dataset) under the admixture model with correlated allele frequencies among populations. Convergence was determined by consistency of likelihood values and cluster assignment between duplicate runs. The optimal number of clusters was determined by calculating the mean ln Pr(X|K) and ΔK values (Evanno *et al.*, 2005) using Structure Harvester (Earl & vonHoldt, 2012). Structure output for the optimal K were combined and summarized using 1000 random repeats of the greedy search algorithm in CLUMPP v. 1.1.2 (Jakobsson & Rosenburg, 2007) on the CLUMPAK server (Kopelman *et al.*, 2015).

Craniometrics

A total of 148 specimens across three taxa—RrC (57), *R. exulans* (61), and *R. hainaldi* (30)—were examined in the craniometric analysis (Table S2, see Louys *et al.*, 2020). We excluded juveniles and very old individuals, as well as those lacking locality information from our sampling. Thirty-eight cranial and mandibular measurements were taken to the nearest 0.01 mm using digital callipers (Mitsutoyo Co.) by the first author following Reutter *et al.* (1999) (basal length (BLL), basilar length (BRL), length of the bullae (BULL), condylobasal length (CBL), condylobasilar length (CBRL), length of the diastema (DA), length of the incisive foramina (FOR), length of the face (GES), length of the braincase (HKL), thickness of the incisor (ID), interorbital breadth (IOB), nasal length (NAS), nasal breadth (NASB), occipital breadth (OCB), occipital length (OCN), length of the upper molar row, alveoli (OZRA), length of the upper molar row, crown (OZRK), palatal length (PL), palatine breadth (PRL), rostral breadth (RB), rostral height (RH), breadth of braincase (SKB), height of braincase with bullae (SKH), and zygomatic breadth (ZYG)) and supplemented with further measurements of the crania (zygomatic plate (ZP), foramen magnum width (FMW), foramen magnum height (FMH), supraoccipital width at the occipital condyles (OCW), and supraoccipital height (BH)); the mandible (mandibular length (ML), mandibular depth (MD), mandibular toothrow length (crown) (MCL), mandibular alveoli length (MAL), thickness lower incisor (LID), minimum corpus length (CL), maximum mandibular height (MDL), mandibular diastema length (LAL), and mandibular depth at M_1 (MID)). These are illustrated on Fig. 1.

Intra-observer error was determined by randomly selecting five skulls for each species and remeasuring them at least a day after first measurement. For each variable, the original and the re-measurement was compared using students t-tests, with the variable rejected if *p* (same mean) < 0.95. Outliers were determined by calculating the 25–75 percent quartile length for each variable for each species, with values more than 3 times this length coded as missing values. In order to summarize the variables and to extract size for further analysis, we subjected the remaining metric variables to a Principal Components Analysis (PCA), using the variance-covariance matrix and disregarding species groupings. PCA was run in PAST v. 2.17c (Hammer *et al.*, 2001). Missing values were treated via pairwise deletion. In PCA application to metric measurements, the first principal component is largely driven by size and can be considered a reasonable body-size proxy.
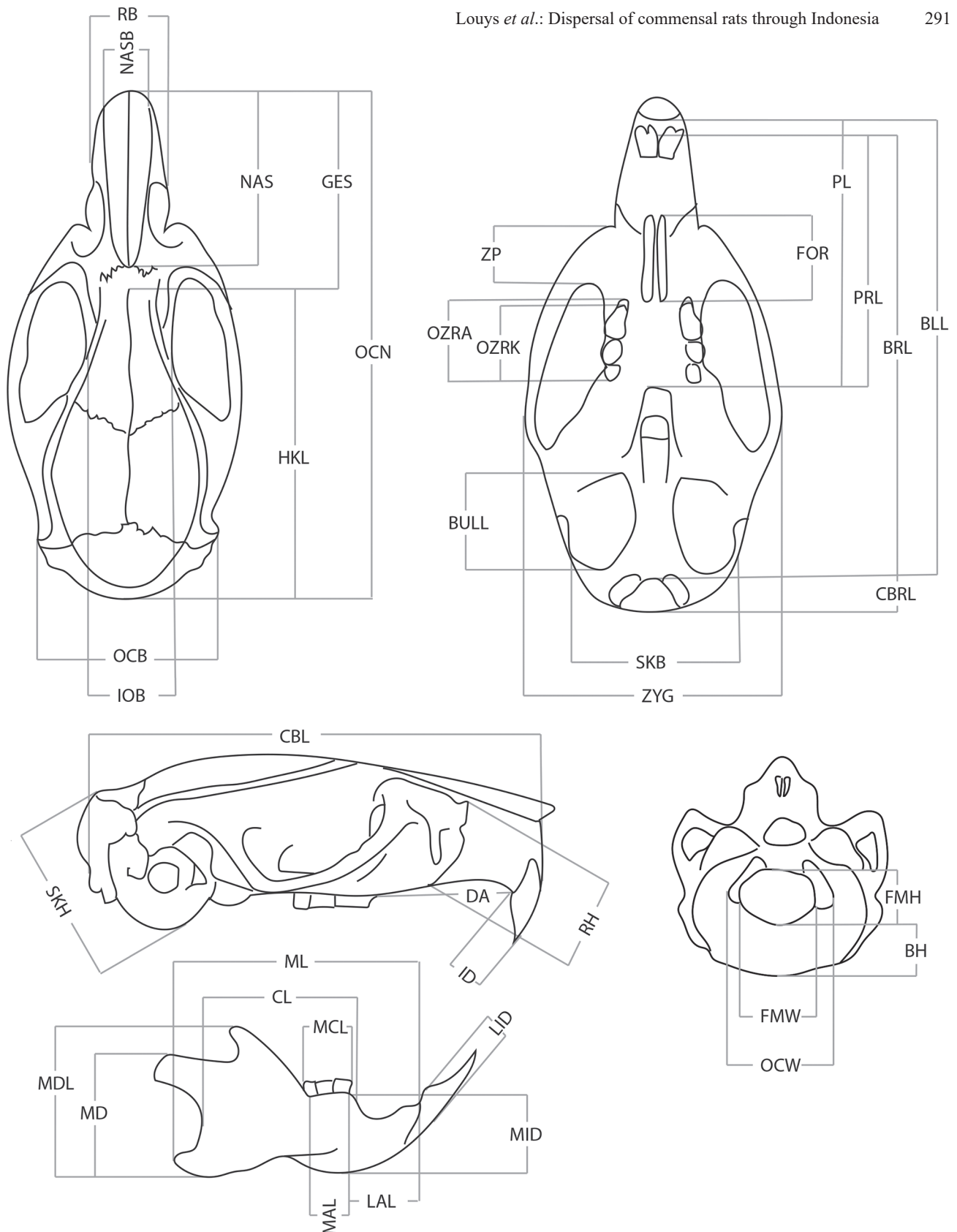
Figure 1. Cranial measurements taken for each specimen: supraoccipital height (BH), basal length (BLL), basilar length (BRL), length of the bullae (BULL), condylobasal length (CBL), condylobasilar length (CBRL), minimum corpus length (CL), length of the diastema (DA), foramen magnum height (FMH), foramen magnum width (FMW), length of the incisive foramina (FOR), length of the face (GES), length of the braincase (HKL), thickness of the incisor (ID), interorbital breadth (IOB), mandibular diastema length (LAL), thickness lower incisor (LID), mandibular alveoli length (MAL), mandibular toothrow length (crown) (MCL), mandibular depth (MD), maximum mandibular height (MDL), mandibular depth at M_1 (MID), mandibular length (ML), nasal length (NAS), nasal breadth (NASB), occipital breadth (OCB), occipital length (OCN), supraoccipital width at the occipital condyles (OCW), length of the upper molar row (alveoli) (OZRA), length of the upper molar row (crown) (OZRK), palatal length (PL), palatine breadth (PRL), rostral breadth (RB), rostral height (RH), breadth of braincase (SKB), height of braincase with bullae (SKH), zygomatic plate (ZP), zygomatic breadth (ZYG).

Figure 2. Sampling locations for skulls included in the craniometric analysis.

Because we were only interested in rats from Nusa Tenggara, we restricted distance analyses to rodents from Adonara, Alor, Flores, Komodo, Lembata, Lombok, Rinca, Roti, Sawu, Semau, Sumba, Sumbawa, and Timor (Fig. 2). We calculated three distance measures corresponding to our three hypotheses of invasion direction: (1) distance from easternmost point (in our dataset, Kalabahi, Alor); (2) distance from Flores (8.7°S and 120.7°E); and (3) distance from westernmost point, (Pelangan, Lombok). Latitude and longitude information provided with each specimen was converted to decimal degrees and distance measures were calculated based on the cartesian distance equation.

Many ecological and environmental pressures can also produce significant changes in body size, particularly for island populations (Efford, 1976; Yom-Tov *et al.*, 1999; Motokawa *et al.*, 2004; Lomolino, 2005; Lomolino *et al.*, 2012; Claude, 2013; van der Geer *et al.*, 2018; Miszkiewicz *et al.*, 2020). These pressures will be unique for each island, and in effect introduce non independence to quantitative data. These can be dampened by averaging values for each island; however, in order to maximize sample size across the entire biogeographical area, a linear mixed model approach that includes each island as a random effect was employed. We modelled the effects of dispersal distance across Nusa Tenggara on rat body size with Bayesian Linear Mixed-Effect Models (BLMEs), with model selection based on Akaike Information Criterion (AIC$_c$) (Burnham & Anderson, 2002). Prior to analysis, all numerical values were centred and scaled. We fitted BLMEs using maximum likelihood parameter estimation criterion with functions implemented in the blme library (Dorie, 2015) of R (R Core Team, 2019). For each analysis, we included island as a random effect and the three distance measures as fixed effects. For the analysis examining all three *Rattus* species, we also included species identity as a fixed effect. To directly compare between models, we standardized input variables with the function available in the arm library (Gelman *et al.*,

2018). We generated the set of models for comparison using the function available in the MuMIn library (Bartoń, 2019) and restricted comparisons to models with at most one fixed distance measure included that fell within two AIC$_c$ of the best model in the set.

Results

Genetics

In the haplotype network for the 66 RrC LIV samples, constructed from a common 376 bp *cyt b* fragment, 20 haplotypes were observed with seven of these found in Nusa Tenggara (Fig. 3). These include Haplotypes 1, 3, 4, 6, 8, 17, and 20, but only H6 is shared between MSEA and Nusa Tenggara. There are two star-shaped clusters originating from H6 and H20. The diversity within H6 "cluster" is found in MSEA and Nusa Tenggara. While the H20 "cluster" is found in the Philippines, Talaud Islands, Sulawesi, and from Lombok and Flores in Nusa Tenggara as well.

In the haplotype network for the 272 RrC LII samples, constructed from a common 812 bp *cyt b* fragment, 100 haplotypes were observed, with seven of these found in Nusa Tenggara (Fig. 4). There are four major clusters in the network, with H1, H12 and H19 forming the focal haplotypes of three of the clusters. These clusters comprise MSEA or East Asian samples only. The fourth cluster, with H6 as the focal haplotype, includes samples from MSEA, the Philippines, and Indonesia. Three further haplotypes, H8 (Sunda), H9(Sunda), H1 (Maluku), were found between the H12 and H19 clusters.

We assessed the relationships among populations of RrC using nuclear microsatellite genotypes at 12 loci (Table S3, see Louys *et al.*, 2020). Genetic cluster analyses did not distinguish members of the RrC from MSEA or in their extralimital range in Indonesia on the basis of their mitochondrial lineage ancestry (Fig. 5). Our findings for

Rattus rattus Complex IV

Rattus exulans
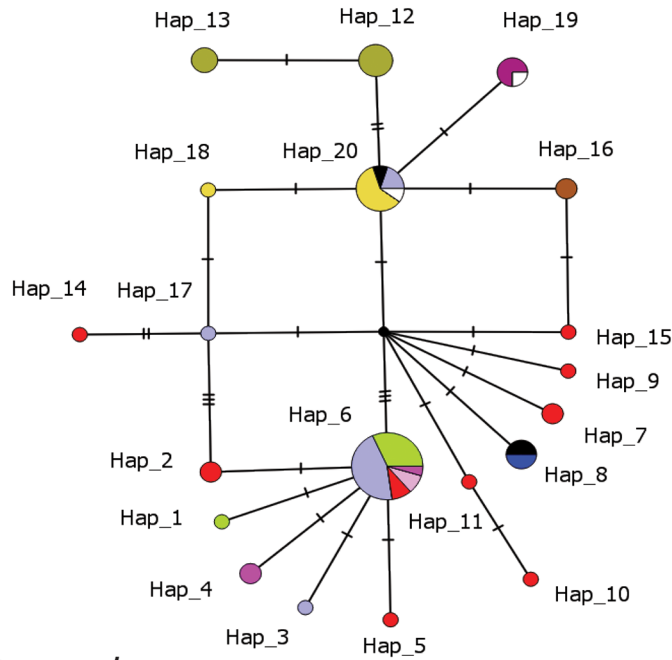
Rattus argentiventer

Nusa Tenggara
- Alor
- Bali
- Flores
- Lombok
- Pantar
- Roti
- Sumba
- Sumbawa
- Timor

Other islands
- Talaud
- Philippines
- New Guinea
- Sulawesi
- Maluku

Regions
- Sunda
- Indonesia
- Pacific
- Mainland Southeast Asia (MSEA)
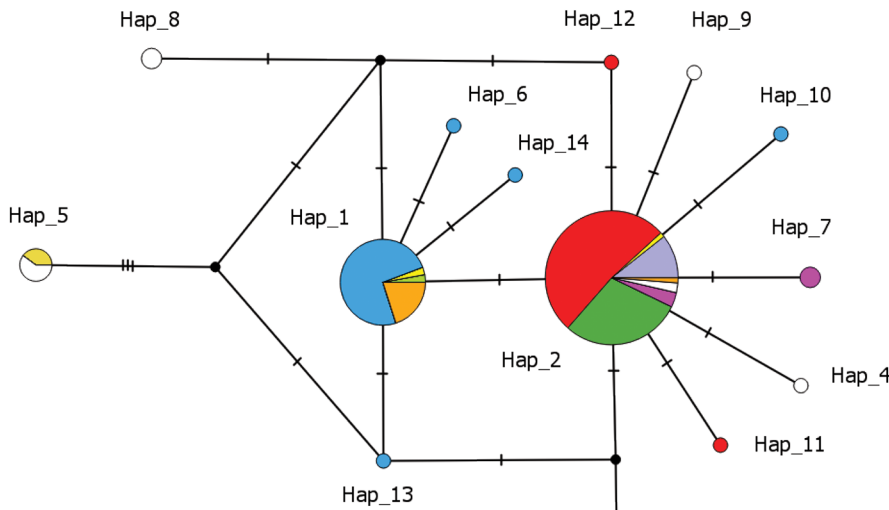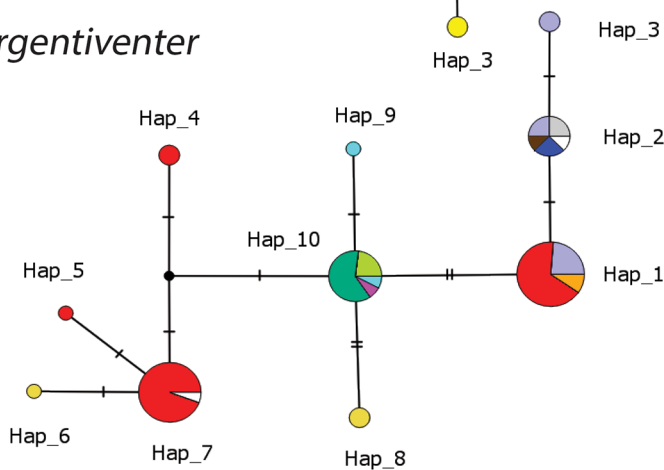
Figure 3. Haplotype networks for *Rattus exulans, R. argentiventer,* and *Rattus rattus* Complex LIV. Sunda refers to the islands of Borneo, Java, and Sumatra; the Indonesian sample (brown) lacks further collection information.
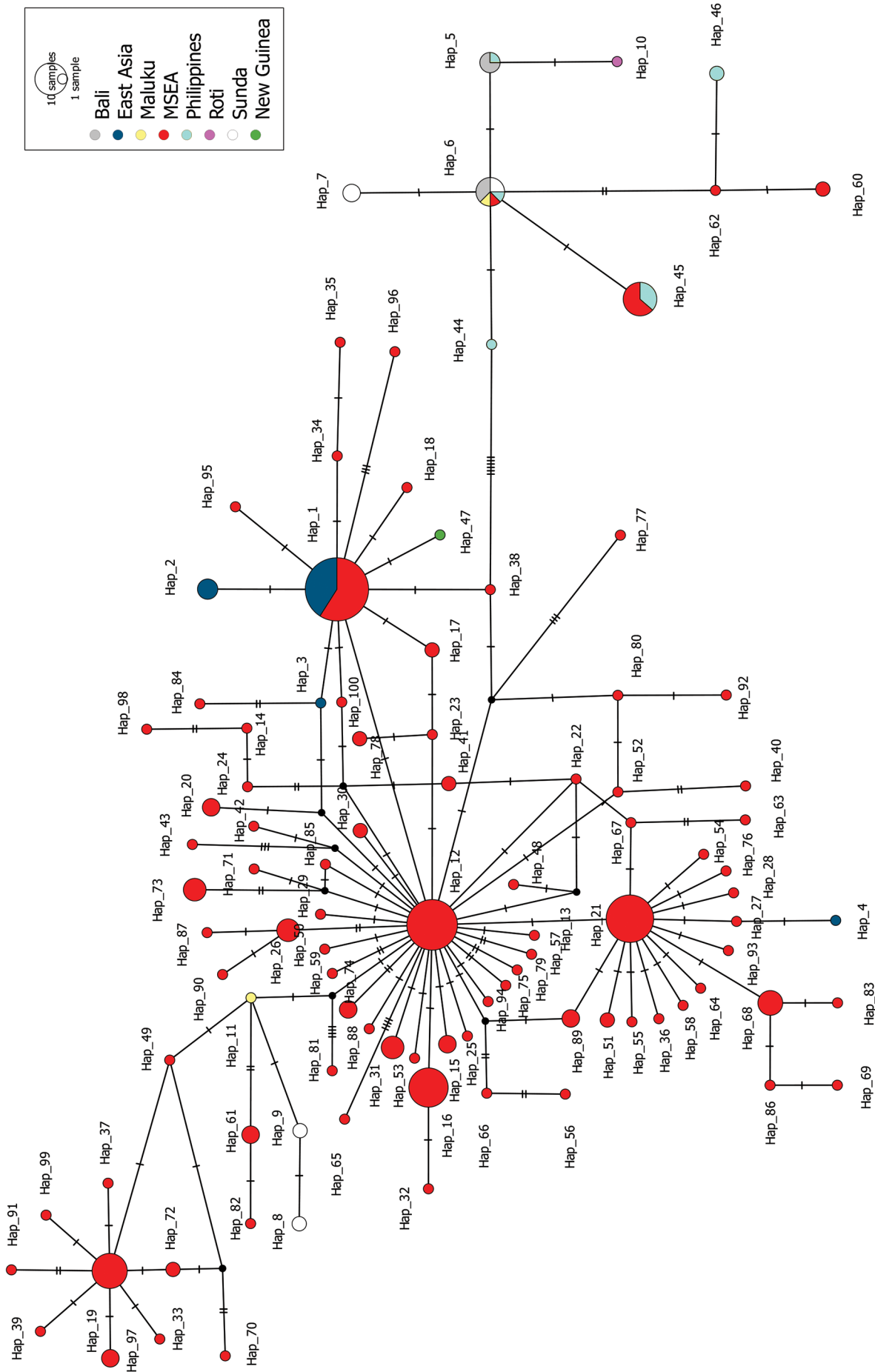
Figure 4. Haplotype network for *Rattus rattus* Complex II. The Nusa Tenggara samples are illustrated on the right of the network.
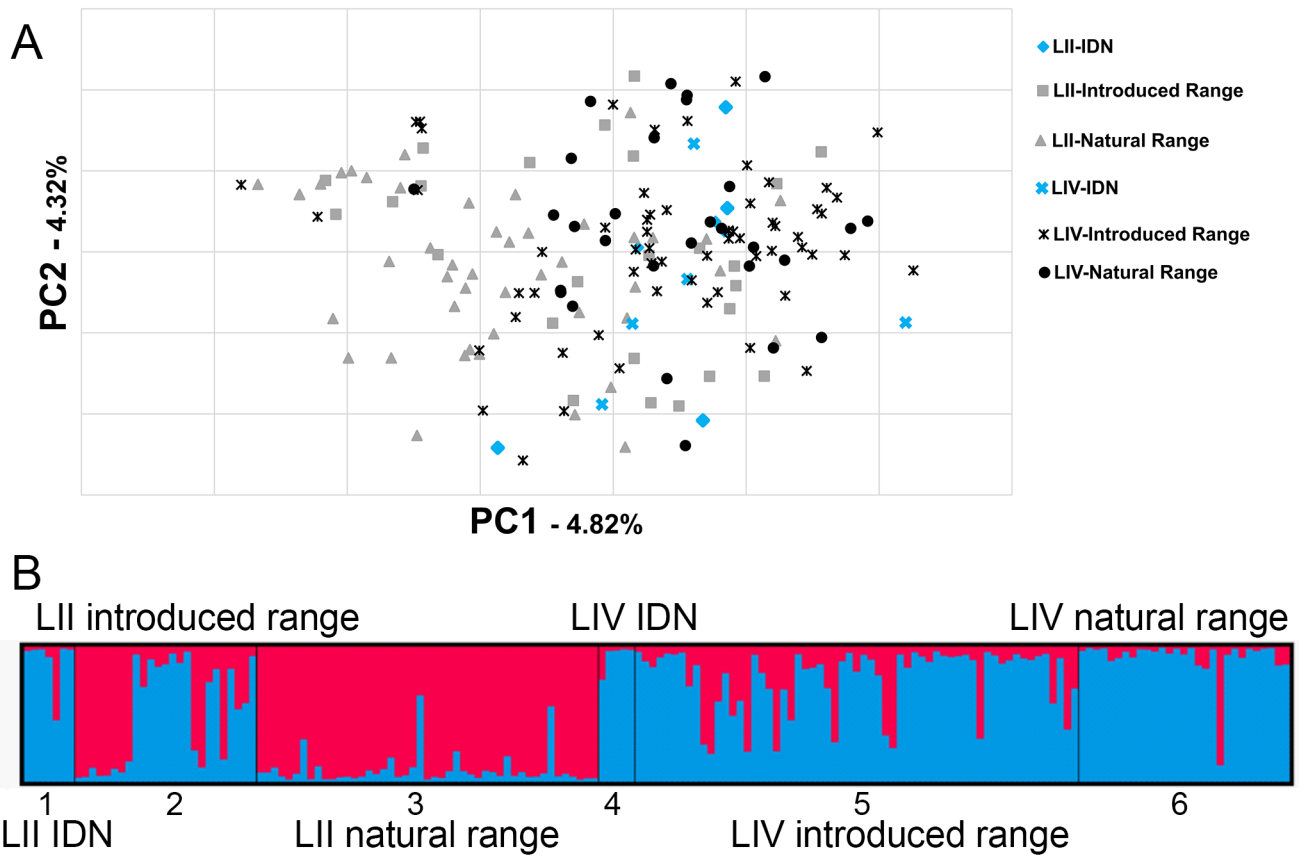
Figure 5. Genetic clustering based on allele frequencies of 12 microsatellite loci genotyped for *Rattus rattus* Complex samples from MSEA and extralimital distribution in Indonesia (IDN). (*A*) PCoA, (*B*) STRUCTURE barplot.

MSEA RrC are consistent with the findings of Pages *et al.* (2013) based on another set of eight microsatellite loci, i.e. the areas sampled for LII and LIV represent broadly introgressed populations.

The *cyt b* network for the 139 *R. exulans* sequences, constructed from the common 330 bp fragment, comprises 14 haplotypes (Fig. 3). Three haplotypes (H2, H11, and H12) were observed from MSEA. However, only H2 is shared with Nusa Tenggara and regions further east such as Maluku and New Guinea. Haplotype 1 is observed in Nusa Tenggara and also in the Pacific, and is the predominant haplotype in the Pacific.

The haplotype network of the *R. argentiventer cyt b* dataset, constructed from the common 312 bp across all 69 individuals, comprises 10 haplotypes (Fig. 3). There were six *cyt b* haplotypes observed in the Nusa Tenggara archipelago (H1, 2, 3, 8, 9, and 10). Haplotype 10 is unique to the region being found in Alor, Pantar, Roti, and Sumba. Only H1 is shared between MSEA and the Nusa Tenggara archipelago, i.e., Flores and Timor. Haplotypes 4, 5, and 7 occur in MSEA, but only H7 is observed in Sundaland (the Sundaic extension of the Asian continental shelf).

Craniometrics

Intra-observer error analysis resulted in the elimination of 19 variables: length diastema (DA), length of the incisive foramina (FOR), length of the face (GES), nasal length (NAS), length of the upper molar row (alveoli) (OZRA), length of the upper molar row (crown) (OZRK), length of the bullae (BULL), occipital breadth (OCB), nasal breadth (NASB), height braincase with bullae (SKH), thickness of incisor (ID), foramen magnum width (FMW), foramen magnum height (FMH), supraoccipital width at the occipital condyles (OCW), supraoccipital height (BH), mandibular depth (MD), mandibular alveoli length (MAL), mandibular diastema length (LAL), and thickness lower incisor (LID). Two values were removed from the dataset following outlier analysis, one from *R. exulans* (M33738: SKB) and another from *R. hainaldi* (M46412: HKL). Univariate statistics of the retained variables are summarized in Table 1. RrC had the largest variance, standard deviation, and coefficient of variation across all craniometric variables. Conversely, *R. exulans* and *R. hainaldi* had very similar values across all these statistics. These two species differ, however, predominately in range and skew, with *R. hainaldi* exhibiting larger ranges and more negative skew in most metrics.

The first principal component (PC1) explained 98% of the variance (Table 2). All craniometric variables contributed positively to PC1, with the largest contributions from condylobasal length (CBL), basal length (BLL), basilar length (BRL), condylobasilar length (CBRL), occipital length (OCN), and mandibular length (ML). Based on log-likelihood values, the "RrC BLME" model performed the best, followed by "*R. exulans*", then the "all rats" model (Table 4). The "all rats" model selection analysis produced two equally likely models. Both included species identity, indicating that species identity significantly impacts models, and implying that they should be treated separately. In the model selection analysis of the RrC only data, three models

Table 2. Univariate statistics for the measurements (mm) used in the craniometric analysis. Abbreviations in caption for Fig. 1. CoV—coefficient of variation.

metric	species	CBL	BLL	BRL	CBRL	OCN	HKL	IOB	PL	PRL	RB	ZYG	RH	SKB	ZP	ML	MCL	CL	MID	MDL
number	*R. rattus*	57	57	57	57	55	57	57	55	55	57	51	56	57	57	57	57	57	57	46
	R. exulans	60	60	60	60	58	60	60	56	56	60	52	60	59	60	58	60	60	60	38
	R. hainaldi	29	29	29	29	29	28	29	28	28	27	29	28	28	27	27	27	27	27	26
mean	*R. rattus*	37.51	35.04	32.13	34.59	39.80	24.06	5.97	21.73	18.80	6.77	18.68	6.52	11.51	4.01	21.33	6.24	13.14	5.79	11.18
	R. exulans	28.70	26.57	24.23	26.34	30.93	19.66	4.79	16.15	13.91	5.21	14.30	4.81	8.92	2.98	15.62	4.85	9.58	4.18	8.30
	R. hainaldi	32.18	29.89	27.26	29.45	34.74	21.36	5.15	18.29	15.74	6.06	16.82	5.76	10.24	3.21	17.95	5.38	10.98	5.07	9.61
minimum	*R. rattus*	30.79	28.28	26.64	29.25	33.25	20.64	4.93	17.76	15.61	4.65	15.93	4.99	8.91	3.01	17.19	5.17	10.78	4.35	8.81
	R. exulans	24.96	22.85	21.04	22.97	27.34	17.52	4.27	14.15	12.25	4.24	12.55	3.80	7.98	2.38	13.53	4.36	7.43	3.65	7.05
	R. hainaldi	28.00	25.31	23.08	25.48	31.01	19.37	4.81	16.58	14.39	5.42	14.88	4.96	9.62	2.70	16.22	4.84	9.82	4.35	8.12
maximum	*R. rattus*	43.66	40.59	37.99	40.92	45.38	26.99	6.88	25.91	22.87	8.30	22.85	8.34	13.66	5.04	26.51	7.35	15.91	7.18	14.09
	R. exulans	31.62	29.74	27.28	29.21	34.49	21.92	5.43	18.15	15.66	5.93	15.45	5.86	10.24	3.42	17.43	5.65	10.66	4.69	9.61
	R. hainaldi	35.03	32.86	30.05	31.91	37.02	23.02	5.48	20.04	17.11	6.62	18.50	6.66	10.97	3.70	19.74	5.91	12.10	5.75	10.90
std. error	*R. rattus*	0.44	0.43	0.40	0.39	0.44	0.20	0.06	0.27	0.24	0.10	0.25	0.12	0.13	0.07	0.31	0.06	0.18	0.09	0.20
	R. exulans	0.20	0.20	0.18	0.18	0.22	0.12	0.03	0.13	0.11	0.05	0.09	0.06	0.05	0.03	0.12	0.03	0.08	0.03	0.08
	R. hainaldi	0.32	0.33	0.29	0.29	0.29	0.16	0.03	0.18	0.14	0.08	0.18	0.07	0.07	0.05	0.18	0.04	0.11	0.06	0.14
variance	*R. rattus*	10.85	10.45	9.04	8.50	10.82	2.36	0.19	4.11	3.20	0.62	3.24	0.77	0.92	0.25	5.38	0.22	1.76	0.45	1.87
	R. exulans	2.34	2.52	2.04	1.94	2.79	0.84	0.05	0.90	0.70	0.13	0.41	0.19	0.17	0.05	0.87	0.06	0.34	0.06	0.26
	R. hainaldi	2.93	3.13	2.47	2.40	2.52	0.72	0.03	0.92	0.59	0.15	0.92	0.15	0.14	0.07	0.92	0.04	0.32	0.10	0.53
standard deviation	*R. rattus*	3.29	3.23	3.01	2.92	3.29	1.54	0.44	2.03	1.79	0.79	1.80	0.88	0.96	0.50	2.32	0.47	1.33	0.67	1.37
	R. exulans	1.53	1.59	1.43	1.39	1.67	0.92	0.22	0.95	0.84	0.36	0.64	0.43	0.42	0.23	0.93	0.25	0.59	0.24	0.51
	R. hainaldi	1.71	1.77	1.57	1.55	1.59	0.85	0.16	0.96	0.77	0.39	0.96	0.39	0.37	0.26	0.96	0.19	0.57	0.32	0.73
25th percentile	*R. rattus*	35.24	32.72	30.1	32.33	37.19	22.94	5.65	20.24	17.43	6.33	17.63	5.92	10.85	3.63	19.51	5.93	12.19	5.41	10.08
	R. exulans	27.86	25.73	23.47	25.64	29.97	19.12	4.64	15.63	13.45	5.00	14.08	4.49	8.62	2.83	15.18	4.69	9.36	4.03	8.04
	R. hainaldi	30.66	28.58	26.23	28.32	33.39	20.81	5.06	17.49	15.18	5.76	15.95	5.47	9.96	3.00	17.48	5.29	10.58	4.89	9.12
75th percentile	*R. rattus*	40.18	37.67	34.68	37.13	43.04	24.94	6.30	23.51	20.30	7.21	19.89	7.32	12.13	4.40	23.12	6.55	14.06	6.45	12.07
	R. exulans	29.66	27.57	25.07	27.40	32.04	20.21	4.92	16.74	14.44	5.50	14.75	5.12	9.24	3.16	16.28	4.95	10.02	4.32	8.56
	R. hainaldi	33.54	31.26	28.45	30.57	35.95	22.11	5.28	18.96	16.42	6.41	17.59	6.04	10.47	3.40	18.69	5.50	11.41	5.30	10.33
skewness	*R. rattus*	0.03	−0.08	0.00	0.11	−0.01	0.22	0.04	0.01	0.05	−0.16	0.40	0.03	−0.27	−0.16	0.20	0.25	0.07	0.04	0.20
	R. exulans	−0.14	−0.20	−0.20	−0.11	0.02	0.41	0.47	0.08	−0.06	−0.10	−0.81	0.09	0.27	−0.24	−0.29	1.09	−1.09	−0.23	0.22
	R. hainaldi	−0.60	−0.77	−0.86	−0.86	−0.57	−0.41	0.07	−0.29	−0.29	−0.06	−0.52	−0.16	0.34	−0.11	−0.20	−0.15	−0.36	−0.02	−0.24
kurtosis	*R. rattus*	−0.80	−0.89	−1.00	−0.80	−0.89	−0.52	−0.42	−0.85	−0.69	0.22	−0.38	−0.98	0.07	−0.79	−0.74	−0.45	−0.61	−0.50	−0.60
	R. exulans	−0.05	0.10	0.11	0.01	−0.04	0.26	0.53	−0.22	−0.29	−0.07	0.54	−0.07	0.66	−0.31	−0.29	1.77	2.06	−0.31	1.15
	R. hainaldi	−0.14	0.35	0.56	0.23	−0.56	0.08	−0.29	−0.75	−0.82	−1.35	−0.57	−0.02	−0.50	−0.66	−0.53	3.24	−0.43	0.00	−0.52
CoV	*R. rattus*	8.78	9.23	9.36	8.43	8.26	6.38	7.29	9.33	9.52	11.65	9.63	13.42	8.34	12.48	10.87	7.58	10.10	11.55	12.24
	R. exulans	5.33	5.97	5.89	5.29	5.40	4.66	4.59	5.86	6.02	6.88	4.49	8.96	4.69	7.61	5.97	5.23	6.11	5.71	6.18
	R. hainaldi	5.32	5.92	5.77	5.26	4.57	3.97	3.16	5.23	4.87	6.45	5.70	6.72	3.61	8.00	5.33	3.51	5.16	6.33	7.59
range	*R. rattus*	4.94	4.95	4.58	4.81	5.85	2.00	0.66	3.27	2.87	0.88	2.26	1.40	1.28	0.77	3.61	0.62	1.87	1.04	1.99
	R. exulans	1.80	1.84	1.60	1.75	2.08	1.09	0.28	1.12	0.99	0.50	0.68	0.63	0.62	0.33	1.10	0.27	0.65	0.29	0.53
	R. hainaldi	2.88	2.68	2.22	2.26	2.56	1.29	0.23	1.47	1.24	0.65	1.64	0.56	0.51	0.40	1.21	0.21	0.83	0.41	1.22

performed equally well. Of these, the top performing model had a negative association between size and distance from the easternmost point. The third model had a positive association between size and distance from westernmost point. These results are consistent in indicating an increase in size moving west to east along Nusa Tenggara. The second model indicated no association between size and distance. The "*R. exulans* only" model selection analysis produced four equally well performing models, with all migration scenarios represented. Thus, for *R. exulans* it appears that size changes of the crania have not been influenced by distance.

Discussion

The haplotype networks of RrC indicate multiple dispersals and back dispersals from mainland Southeast Asia (MSEA) into Wallacea. The RrC LII network indicates an early dispersal into the Philippines, and hence into Maluku and Bali, judging by the significant distance between mainland Southeast Asian haplotype group 38 and the Philippine haplotype group 44. Only one RrC LII haplotype group (H10) is recovered from Nusa Tenggara, which stems off from a haplogroup found in Bali and the Philippines. In RrC LIV, Haplotype group 6, which is represented in Alor, Flores, Pantar, Roti, and MSEA, shows considerable distance from the reconstructed core area indicating a single dispersal event followed by *in situ* production of genetic diversity (Fig. 3). This dispersal is thus also likely to be of considerable antiquity; however, it cannot be determined whether genetic diversity emerged mainly in MSEA with dispersal into Nusa Tenggara, or if dispersal was to Nusa Tenggara, with subsequent back-dispersal into MSEA. Given the additional diverging haplotypes from Haplotype group 6 into both MSEA and Alor, Roti, and Flores, considerable genetic exchange occurred during this period. A second, more recent dispersal of RrC through Southeast Asia is indicated by the fanning pattern of network connections with fewer haplotype substitutions from the reconstructed core. This dispersal was multidirectional, with several unique haplotypes now found in MSEA, Flores, and Sumbawa-Lombok, Philippines-Sundaland-Lombok-Flores. The latter haplogroup gave rise to the Sulawesi and Talaud RrC LIV

Table 3. Summary of Principal Components Analysis. Eigenvalues and amount of variance explained by each listed in columns on the left; loading of each cranial variable for the first two principal component (PC) scores on the right. Abbreviations in caption for Fig. 1.

PC	eigenvalue	% variance	cranial variable	PC1	PC2
1	143.444	98.015	CBL	0.387	−0.162
2	0.591	0.404	BLL	0.375	−0.276
3	0.553	0.378	BRL	0.348	−0.168
4	0.392	0.268	CBRL	0.357	−0.043
5	0.263	0.180	OCN	0.389	0.091
6	0.191	0.130	HKL	0.188	0.359
7	0.142	0.097	IOB	0.047	0.034
8	0.109	0.074	PL	0.242	−0.116
9	0.107	0.073	PRL	0.212	0.009
10	0.074	0.051	RB	0.071	0.045
11	0.061	0.041	ZYG	0.187	0.455
12	0.059	0.041	RH	0.077	−0.120
13	0.051	0.035	SKB	0.108	0.203
14	0.040	0.027	ZP	0.046	0.014
15	0.032	0.022	ML	0.252	0.376
16	0.027	0.019	MCL	0.054	0.135
17	0.014	0.010	CL	0.154	0.100
18	0.008	0.006	MID	0.069	0.089
19	−0.190	−0.130	MDL	0.123	−0.523

haplogroups, suggested by Louys *et al.* (2018a) to belong to the Austronesian expansion of agriculturalists (Bellwood, 2007). This dispersal is likely tied to the one observed between Bali-Philippines and Roti in RrC LII. While introgression in Nusa Tenggara between RrC LII and RrC LIV may have occurred during this later dispersal, MSEA and other extralimital populations, e.g., the Philippines that comprise likely sources for dispersal into Indonesia, have highly admixed nuclear genomes.

The mtDNA history likely reflects dispersal episodes and sources but not the overall biological phenotype of rats arriving in Indonesia. Thus, we feel justified in treating the Indonesian material in the craniometric analysis as being derived from a single biological species. The name of this taxon remains uncertain. Robins *et al.* (2007) associated *R. rattus diardii*, the Malayan house rat, with RrC LIV and Musser & Carlton (2005) list *R. rattus diardii* as synonymous with *R. tanezumi*, which in turn was considered part of RrC LIV by Denys *et al.* (2017). At present, the populations

of LII and LIV mitochondrial ancestry rats are perhaps best viewed as "the *Rattus tanezumi* Complex" within the RrC. We determined the mitochondrial haplotype for eight individuals, six were LIV and two were LII (Table S2, see Louys *et al.*, 2020).

Our craniometric analysis provides further support for an earlier dispersal of RrC into Nusa Tenggara. The traits under evolutionary selection for species at the population front are likely multifaceted and would not necessarily be restricted to phenotypic changes (Chuang & Peterson, 2016). Nevertheless, in this instance the craniometric analysis indicates an increase in the length of the skull associated with an eastern movement through Nusa Tenggara (Fig. 6). Increases in skull length are strongly correlated with increased body mass in rodents (Bertrand *et al.*, 2016). Thus, body size of these rats increased through the archipelago, consistent with observations made for a range of demes at the population edge relative to the core (Chuang & Peterson, 2016). This is inferred to be due to the physical exertion and

Table 4. Bayesian Linear Mixed-Effect Model summary statistics and Akaike Information Criterion (AICc) comparisons. *RrC, Rattus rattus* Complex; *z-D(east)* is the distance from the east fixed effect, *z-D(Flores)* is the distance from Flores fixed effect, and *z-D(west)* is the distance from the west fixed effect.

model	intercept	species	z-D(east)	z-D(Flores)	z-D(west)	df	logLn	AIC$_c$	delta	weight
all rats 1	−0.82570	+	—	—	—	5	−93.096	196.7	0	0.418
all rats 2	−0.82740	+	−0.11950	—	—	6	−92.756	198.2	1.52	0.195
R. exulans 1	−0.06516	—	—	—	—	3	−82.830	172.1	0	0.259
R. exulans 2	−0.02467	—	0.6714	—	—	4	−81.927	172.6	0.49	0.203
R. exulans 3	−0.06115	—	—	0.5819	—	4	−82.350	173.4	1.34	0.133
R. exulans 4	−0.03825	—	—	—	−0.4393	4	−82.351	173.4	1.34	0.133
RrC 1	−0.04494	—	−0.6587	—	—	4	−54.671	118.5	0	0.285
RrC 2	−0.02832	—	—	—	—	3	−56.248	119.2	0.68	0.203
RrC 3	−0.03502	—	—	—	0.49850	4	−55.263	119.7	1.18	0.158
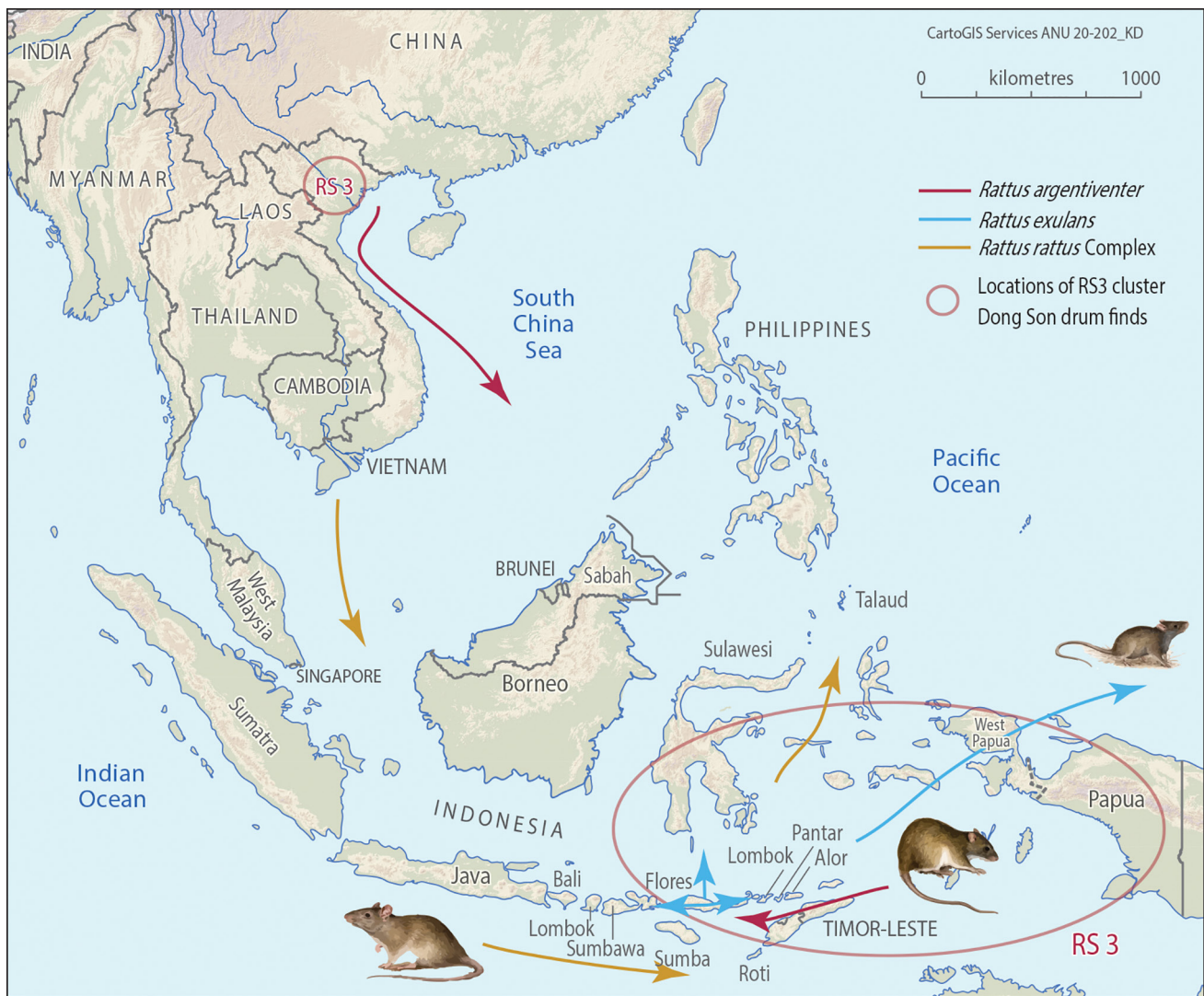
Figure 6. Summary of likely movements of the three commensal rodents through the Nusa Tenggara island chain. Rat illustrations redrawn and adapted from R. Budden, location of RS3 Dong Song drums from Calo (2014).

activity required by active dispersal. It suggests that at least one of the dispersals of RrC into Nusa Tenggara was an active one by the species as opposed to a passive translocation (e.g., as part of a Neolithic "package"). This active invasion could correspond to the more ancient dispersal event suggested by the haplotype network.

The *R. exulans* haplotype network largely mirrors that reported by Thomson *et al.* (2014) (see also Matisoo-Smith *et al.*, 2014; Hingston, 2015; West *et al.*, 2017). It shows a major haplotype group (centred on haplotype 2) with representatives currently found on MSEA, New Guinea, Roti, Sundaland, Timor, Flores, and the Maluku Islands, and a second major group (centred on haplotype 1) representing a major dispersal into the Pacific (Fig. 3). New sequences added by this study include an individual from Alor, a fossil specimen from Liang Luar cave, Flores, five specimens from Roti, and 14 specimens from archaeological deposits in Papua New Guinea (Panakiwuk, New Ireland; and Paleflatu, Sandaun Province). The Alor specimen belongs to the Pacific haplogroup, while the Liang Luar specimen corresponds to haplogroup 2, as was found by Thomson *et al.* (2014). All the specimens from Roti and New Guinea are also associated

with this haplogroup, which is considered to include the source population of *R. exulans*. This was suggested to be Flores by Schwartz & Schwartz (1967) based on the unique occurrence of the wild type white-belly phenotype on Flores and supported by Thomson *et al*. (2014) who observed the highest level of genetic diversity on that island. Our results do not contradict this conclusion. Archaeological records from Timor record the introduction of at least *R. exulans* on that island several thousand years before present (Glover, 1986), suggesting an initial Nusa Tenggara dispersal potentially co-eval or soon after the dispersal of RrC through the islands.

Unlike the *R. rattus* craniometric analysis, no one model of size change associated with distance was favoured by our *R. exulans* analysis. This was unexpected, as based on the dispersal of *R. exulans* out of Flores and eventually into the Pacific, we expected size to vary as a measure of distance from this island. It appears that size does not correlate with direction of travel throughout Nusa Tenggara, and in fact craniometric values for this species show similar statistics to the Flores-restricted endemic *R. hainaldi*. We interpret this result to indicate passive dispersal by this murid. Traits selected for at the expanding edge of this species, at least

in Nusa Tenggara, are not discernible from their cranial measurements, and are instead more consistent with *R. exulans* belonging to a commensal Neolithic "package". Such a dispersal mechanism would more likely favour traits associated with higher neophilia (Chuang & Peterson, 2016) and domestication (Leppard, 2018).

The zoogeographic history of *R. argentiventer* is largely unknown. Aplin *et al.* (2003) suggested a wetland or grassland origin for *R. argentiventer* based on its physiological link to the tillering stage of the rice crop. Its widespread distribution is attributed to human activities and unintentional transportation (Harrison, 1961). Musser (1973) suggested that the Sulawesi, Philippine, and New Guinea populations were likely recent anthropogenic introductions and remarked on its morphological uniformity throughout its range. Maryanto (2003), however, was able to distinguish four groups based on a craniometric analysis: a Java group, a Bali-Sulawesi group, a Sumba group, and a Kalimantan group. Maryanto (2003) agreed with Musser's (1973) conclusion that the Sulawesi population was likely a recent introduction. Our haplotype network analysis indicates that the Philippines populations were likely recent introductions from MSEA and Nusa Tenggara. Our haplotype groups 9–10 probably correspond to the Sumba group of Maryanto (2003), the single specimen from Kalimantan that we sampled belongs to haplogroup 7, a group otherwise represented in MSEA. Our single specimen from Sumatra belongs to haplogroup 2. Maryanto (2003) could not distinguish between Java, Sumatra, and Thailand specimens, and these populations are likely represented by our haplogroups 1–3. Interestingly, these also include specimens from Liang Luar cave in Flores. These are not represented in the better sampled Liang Bua deposits of Flores until the mid to late Holocene (Locatelli *et al.*, 2015), indicating that haplogroups 1–3 represents a Neolithic or Metal Age dispersal. This cluster is separated from the "Sumba" group—actually representing a Nusa Tenggara Timur cluster—by two substitutions and is isolated from haplogroup 7. Because of this, we suggest that *R. argentiventer* dispersed directly from MSEA into eastern Nusa Tenggara, followed by a back dispersal into MSEA following a residence in Nusa Tenggara of sufficient duration to produce the haplotype diversity observed (Fig. 6).

Interestingly, the above model for the introduction of *R. argentiventer* matches well with the timing and distribution of the Metal Age maritime dispersal of Dong Son drums from mainland Southeast Asia into Nusa Tenggara. Dong Son drums (also known as Heger I drums) are large and elaborately decorated cast bronze drums made by the Dong Son culture of northern Vietnam/southern China in the last few centuries BC (Heger, 1902; Calo, 2014). These high prestige objects were traded along river and sea routes, including subsequent entry into maritime networks, and are found throughout MSEA and ISEA (Calo, 2014). A detailed study of Dong Son drums by Calo (2014) used the stylistic features and distributions of around 400 drums in Southeast Asia to derive distinct drum groups and trading clusters. One such group was found to characterize an eastern Indonesian dispersal (known as the RS3 cluster) whose homogeneity indicates a short and intensive distribution. RS3 drums have been found in many islands of the Nusa Tenggara archipelago and Maluku including Sangeang, Selayar, Alor, Roti, Timor, Leti, Luang, Tanimbar, Kei, Serua, Gorom, and Buru, and in western New Guinea (Calo, 2014; Oliviera *et al.*, 2019).

A two-step process for the dispersal of the RS3 cluster was suggested, with drums initially moving from northern Vietnamese and Chinese production centres to the islands, via a direct maritime route, approximately 1850–1600 years ago (Calo, 2014). Inter-island maritime networks moved them into Nusa Tenggara and thence eastwards (Calo, 2014). This is thought to have begun after the fifth century AD and may have continued as late as the first millennium AD (c. 1600 years ago through to 1000 years ago).

The maritime networks that facilitated the movement of Dong Son drums into Nusa Tenggara may have had the unintentional consequence of also introducing *R. argentiventer* into the island chain. The establishment of this rodent species in the islands may not have been successful prior to this point owing to a lack of suitable habitat, specifically cultivated fields. Although it is currently unknown when rice cultivation in Nusa Tenggara began, Bellwood (2011) suggested the third phase of wet rice cultivation from China into Southeast Asia occurred after 500 BC in regions of high population growth such as Java and Bali. However, in drier areas unsuitable for intensive wet rice cultivation, such as the Nusa Tenggara islands east of the Wallace Line, this may have occurred later, coincident with the Dong Son drum expansion detailed above. Importantly, the latter is also coincident with the extinction times of giant endemic rats from Sumba, Timor, and Alor, which have been associated with the widespread introduction of metal tools into Nusa Tenggara facilitating large-scale forest clearance (O'Connor & Aplin, 2007; Louys *et al.*, 2018b; Miszkiewicz *et al.*, 2020). A later invasion into the western part of the island chain is suggested by the fossil record of Liang Luar, a natural Holocene assemblage from Flores, described by St Pierre (2011). In this sequence, invasive rodents including *R. argentiventer* and *R. rattus* do not appear until approximately the last 400 years (*Rattus exulans* is present from at least 2500 BP). St Pierre (2011) suggests that this rodent record is tied with the introduction of wet rice agriculture into Flores, possibly from Sumbawa or Sulawesi and temporally associated with changes in local governments and land use practices. However, our haplotype network (Fig. 3) suggest the Liang Luar haplotypes branch off from eastern Nusa Tenggara populations, thus potentially pre-dating these changes. It is possible the Liang Luar deposit records a local introduction event only. Sifting through these different scenarios will require more intensive sampling of the late Holocene rodent record of these islands, as well as more intensive genetic sampling.

Conclusions

Our analyses suggest the invasion of Nusa Tenggara by *Rattus* species occurred at different times and from different directions. The first introduction was likely by the *Rattus rattus* Complex from mainland Southeast Asia into the western parts of Nusa Tenggara and moving eastwards. Phenotypic changes associated with this migration include an increase in skull length, suggesting classic invasion effects selecting for larger and better-conditioned demes at the invasion front relative to the core. The movement of *Rattus exulans* from Flores is not associated with any craniometric changes in either direction, or from Flores as a point of origin. Instead, it is possible that this invasion

favoured opportunistic individuals with a propensity for domestication. This dispersal may have occurred coeval with, or perhaps after the introduction of the *Rattus rattus* Complex into Nusa Tenggara. The invasion of *R. argentiventer* into Nusa Tenggara was likely the last of the three introductions considered here. Genetic analysis in combination with a consideration of major maritime exchange networks operating at the time suggests that it proceeded from the eastern parts of Nusa Tenggara westwards, required the establishment by people of open grasslands and cultivated fields starting from approximately 1500 years ago, and only established a population in the western part of the archipelago in the last 400 years.

Supplementary data

Tables S1, S2 and S3 are published separately by the authors (see Louys *et al.*, 2020).

Table S1. Sample information for murids included in the haplotype network analyses. PCR primers used for contemporary samples of *Rattus exulans* and *R. argentiventer*—A835/A1937; and for ancient samples—A1936/A1937 (Louys *et al.*, 2020: table S1).
https://doi.org/10.6084/m9.figshare.12996851

Table S2. Murid crania data, raw measurements (mm). All specimens are registered in the Western Australian Museum. The mitochondrial lineage of eight RrC sequenced are indicated next to their registration number (Louys *et al.*, 2020: table S2).
https://doi.org/10.6084/m9.figshare.12996851

Table S3. Nuclear microsatellite genotypes for RrC samples.
https://doi.org/10.6084/m9.figshare.12996851

References

Aplin, K. P., T. Chesser, and J. ten Have. 2003. Evolutionary biology of the genus *Rattus*: profile of an archetypal rodent pest. In *Rats, Mice and People: Rodent Biology and Management*, ed. G. R. Singleton, L. A. Hinds, C. J. Krebs, and D. M. Spratt, pp. 487–498. ACIAR Monograph no. 96, 564 pp.

Aplin, K. P., H. Suzuki, A. A. Chinen, R. T. Chesser, J. ten Have, S. C. Donnellan, J. Austin, A. Frost, J. P. Gonzalez, V. Herbreteau, F. Catzeflis, J. Soubrier, Y.-P. Fang, J. Robins, E. Matisoo-Smith, A. D. S. Bastos, I. Maryanto, M. H. Sinaga, C. Denys, R. A. Van Den Bussche, C. Conroy, K. Rowe, and A. Cooper. 2011. Multiple geographic origins of commensalism and complex dispersal history of Black Rats. *PLoS ONE* 6(11): e26357.
https://doi.org/10.1371/journal.pone.0026357

Bartoń, K. 2019. *MuMIn: Multi-Model Inference. R package*. Version 1.43.10.
https://cran.r-project.org/web/packages/MuMIn/index.html

Bellwood, P. 2007. *Prehistory of the Indo-Malaysian Archipelago*, revised edition. Canberra: ANU E Press.
https://doi.org/10.26530/OAPEN_459472

Bellwood, P., 2011. The checkered prehistory of rice movement southwards as a domesticated cereal from the Yangzi to the equator. *Rice* 4: 93.
https://doi.org/10.1007/s12284-011-9068-9

Bertrand, O. C., M. A. Schillaci, and M. T. Silcox. 2016. Cranial dimensions as estimators of body mass and locomotor habits in extant and fossil rodents. *Journal of Vertebrate Paleontology* 36: e1014905.
https://doi.org/10.1080/02724634.2015.1014905

Burnham, K. P., and D. R. Anderson. 2002. *Model Selection and Multimodel Inference: a Practical Information-Theoretic Approach. Ecological Modelling*. New York: Springer Science & Business Media.

Calo, A. 2014. *Trails of Bronze Drums Across Early Southeast Asia: Exchange Routes and Connected Cultural Spheres*. Singapore: ISEAS.

Chuang, A., and C. R. Peterson. 2016. Expanding population edges: theories, traits, and trade-offs. *Global Change Biology* 22: 494–512.
https://doi.org/10.1111/gcb.13107

Claude, J. 2013. Log-shape ratios, Procrustes superimposition, elliptic Fourier analysis: three worked examples in R. *Hystrix, the Italian Journal of Mammalogy* 24(1): 94–102.
https://doi.org/10.4404/hystrix-24.1-6316

Denys, C., P. J. Taylor, and K. P. Aplin. 2017. Family Muridae (true mice and rats, gerbils and relatives). In *Handbook of the Mammals of the World. Vol. 7 Rodents II*, ed. D. E. Wilson, T. E. Lacher Jr, and R. A. Mittermeier, pp. 536–886. Barcelona: Lynx Edicions.

Dorie, V. 2015. blme: Bayesian Linear Mixed-Effects Models. R package. Version 1.0-4.
https://cran.r-project.org/web/packages/blme/index.html

Earl, D. A., and B. M. Vonholdt. 2012. Structure harvester: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* 4: 359–361.
https://doi.org/10.1007/s12686-011-9548-7

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

Efford, M. 1976. Rattus exulans *in Polynesia—a Case of Morphometric Divergence*. Unpublished B.Sc. (Hons) thesis. Victoria University, Wellington, New Zealand.

Evanno, G., S. Regnaut, and J. Goudet. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* 8: 2611–2620.
https://doi.org/10.1111/j.1365-294X.2005.02553.x

Faircloth, B. 2008. MSATCOMMANDER: detection of micro-satellite repeat arrays and automated, locus-specific primer design. *Molecular Ecology Resources* 8: 92–94.
https://doi.org/10.1111/j.1471-8286.2007.01884.x

Gardner, M. G., A. Fitch, T. Bertozzi, and A. Lowe. 2011. Rise of the machines—recommendations for ecologists when using next generation sequencing for microsatellite development. *Molecular Ecology Resources* 11: 1093–1101.
https://doi.org/10.1111/j.1755-0998.2011.03037.x

Gelman, A., Y.-S. Su, M. Yajima, J. Hill, M. Grazia Pittau, J. Kerman, T. Zheng, and V. Dorie, V. 2018. *arm*: data analysis using regression and multilevel/hierarchical models. R package. Version 1.10-1.
https://cran.r-project.org/web/packages/arm/index.html

Glover, I. 1986. *Archaeology in eastern Timor, 1966–67. Terra Australis* 11. Canberra: Department of Prehistory, Research School of Pacific Studies, The Australian National University.

Hammer, Ø., D. A. T. Harper, and P. D. Ryan. 2001. PAST: paleontological statistics software package for education and data analysis. *Palaeontologia Electronica* 4(1), 9 pp.

Harrison, J. L. 1961. Ecology of the forms of *Rattus rattus* in the Malay Peninsula. *Proceedings of the Ninth Pacific Science Congress* 19: 19–24

Hayden, M. J., T. M. Nguyen, A. Waterman, and K. J. Chalmers. 2008. Multiplex-ready PCR: a new method for multiplexed SSR and SNP genotyping. *BMC Genomics* 9: 80.
https://doi.org/10.1186/1471-2164-9-80

Heger, F. 1902. *Alte metalltrommeln aus Südost-Asien*. Leipzig: Kommissions-Verlag von Karl W. Hiersemann.

Hingston, M. 2015. *Phylogeography of the Commensal* Rattus exulans *with Implications for its use as Bioproxy for Human Migrations*. Unpublished Ph.D. thesis. University of Auckland.

Jakobsson, M., and N. A. Rosenberg. 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 23: 1801–1806.
https://doi.org/10.1093/bioinformatics/btm233

Kitchener, D. J., R. A. How, and Maharadatunkamsi. 1991. A new species of *Rattus* from the mountains of West Flores, Indonesia. *Records of the Western Australian Museum* 15: 611–626.

Kopelman, N. M., J. Mayzel, M. Jakobsson, N. A. Rosenberg, and I. Mayrose. 2015. CLUMPAK: a program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources* 15: 1179–1191.
https://doi.org/10.1111/1755-0998.12387

Lack, J. B., D. U. Greene, C. J. Conroy, M. J. Hamilton, J. K. Braun, M. A. Mare, and R. A. Van Den Bussche. 2012. Invasion facilitates hybridization with introgression in the *Rattus rattus* species complex. *Molecular Ecology* 21: 3545–3561.
https://doi.org/10.1111/j.1365-294X.2012.05620.x

Leigh, J. W., and D. Bryant. 2015. Popart: full-feature software for haplotype network construction. *Methods in Ecology and Evolution* 6: 1110–1116.
https://doi.org/10.1111/2041-210X.12410

Leppard, T. P. 2018. Rehearsing the Anthropocene in micro-cosm: the palaeoenvironmental impacts of the Pacific rat (*Rattus exulans*) and other non-human species during island Neolithization. In *Multispecies Archaeology,* ed. S. E. Pilaar Birch, pp. 47–64. Routledge.
https://doi.org/10.4324/9781315707709-4

Locatelli, E., R. Due, and L. W. van den Hoek Ostende. 2015. Middle-sized murids from Liang Bua (Flores, Indonesia): insular endemics, human introductions and palaeoenvironment. *Palaeobiodiversity and Palaeoenvironments* 95: 497–512.
https://doi.org/10.1007/s12549-015-0204-1

Lomolino, M. V. 1985. Body size of mammals on islands: the island rule re-examined. *American Naturalist* 125: 310–316.
https://doi.org/10.1086/284343

Lomolino, M. V., D. F. Sax, P. R. Palombo, and A. A. van der Geer. 2012. Of mice and mammoths: evaluations of causal explanations for body size evolution in insular mammals. *Journal of Biogeography* 39: 842–854.
https://doi.org/10.1111/j.1365-2699.2011.02656.x

Louys, J., M. Herrera, S. Hawkins, K. Aplin, C. Reepmeyer, F. Hopf, S. C. Donnellan, S. O'Connor, and D. A. Tanudirjo. 2018a. Neolithic dispersal implications of murids from late Holocene archaeological and modern natural deposits in the Talaud Islands, northern Sulawesi. In *The Archaeology of Sulawesi: Current Research on the Pleistocene to the Historic Period*, ed. S. O'Connor, D. Bulbeck, and J. Meyer, pp. 223–242, *Terra Australis* 48. Canberra: ANU E Press.
https://doi.org/10.22459/TA48.11.2018.14

Louys, Julien, Michael B. Herrera, Vicki A. Thomson, Andrew S. Wiewel, Stephen C. Donnellan, Sue O'Connor, and Ken Aplin. 2020. Supplementary data for craniometric and genetic analyses of commensal rats in Nusa Tenggara, Indonesia. *figshare.* Dataset.
https://doi.org/10.6084/m9.figshare.12996851

Louys, J., S. O'Connor, Mahirta, P. Higgins, S. Hawkins, and T. Maloney. 2018b. New genus and species of giant rat from Alor Island, Indonesia. *Journal of Asia-Pacific Biodiversity* 11: 503–510.
https://doi.org/10.1016/j.japb.2018.08.005

Maryanto, I. 2003. Taxonomic status of the ricefield rat *Rattus argentiventer* (Robinson and Kloss, 1916) (Rodentia) from Thailand, Malaysia and Indonesia based on morphological variation. *Records of the Western Australian Museum* 22: 47–66.
https://doi.org/10.18195/issn.0312-3162.22(1).2003.047-065

Matisoo-Smith, E., J. H. Robins, and R. C. Green. 2004. Origins and dispersals of Pacific peoples: evidence from mtDNA phylogenies of the Pacific rat. *Proceedings of the National Academy of Sciences USA* 101: 9167–9172.
https://doi.org/10.1073/pnas.0403120101

Meglécz, E. 2007. MicroFamily: a computer program for detecting flanking region similarities among different microsatellite loci. *Molecular Ecology Notes* 7: 2611–2620.
https://doi.org/10.1111/j.1471-8286.2006.01537.x

Miszkiewicz, J. J., J. Louys, R. M. D. Beck, P. Mahoney, K. Aplin, and S. O'Connor. 2020. Island rule and bone metabolism in fossil murines from Timor. *Biological Journal of the Linnean Society* 129(3): 570–586.
https://doi.org/10.1093/biolinnean/blz197

Motokawa, M., L.-K. Lin, and K.-H. Lu. 2004. Geographic variation in cranial features of the Polynesian rat *Rattus exulans* (Peale, 1848) (Mammalia: Rodentia: Muridae). *The Raffles Bulletin of Zoology* 52(2): 653–663.

Musser, G. G. 1973. Zoogeographical significance of the ricefield rat *Rattus argentiventer* on Celebes and New Guinea and the identity of *Rattus pesticulus*. *American Museum Novitates* 2511: 1–30

Musser, G. G., and M. D. Carleton. 2005. Family Muridae. In *Mammal Species of the World: a Taxonomic and Geographic Reference*, ed. D. E. Wilson and D. M. Reeder, pp. 894–1531. Baltimore: The Johns Hopkins University Press.

Nicholls, J. A., and J. J. Austin. 2005. Phylogeography of an east Australian wet-forest bird, the satin bowerbird (*Ptilonorhynchus violaceus*), derived from mtDNA, and its relationship to morphology. *Molecular Ecology* 14: 1485–1496.
https://doi.org/10.1111/j.1365-294X.2005.02544.x

O'Connor, S., and K. Aplin. 2007. A matter of balance: an overview of Pleistocene occupation history and the impact of the Last Glacial Phase in East Timor and the Aru Islands, eastern Indonesia. *Archaeology in Oceania* 42(3): 82–90.
https://doi.org/10.1002/j.1834-4453.2007.tb00021.x

Oliveira, N., S. O'Connor, and P. Bellwood. 2019. Dong Son drums from Timor-Leste: prehistoric bronze artefacts in Island Southeast Asia. *Antiquity* 93: 163–180.
https://doi.org/10.15184/aqy.2018.177

Pages, M., E. Bazin, M. Galan, Y. Chaval, J. Claude, V. Herbreteau, J. Michaux, S. Piry, S. Morand, and J.-F. Cosson. 2013. Cytonuclear discordance among Southeast Asian black rats (*Rattus rattus* complex). *Molecular Ecology* 22: 1019–1034.
https://doi.org/10.1111/mec.12149

Peakall, R., and P. E. Smouse. 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research- an update. *Bioinformatics* 28: 2537–2539.
https://doi.org/10.1093/bioinformatics/bts460

Phillips, B. L., G. P. Brown, J. K. Webb, and R. Shine. 2006. Invasion and the evolution of speed in toads. *Nature* 439: 803.
https://doi.org/10.1038/439803a

Phillips, B. L., G. P. Brown, M. Greenlees, J. K. Webb, and R. Shine. 2007. Rapid expansion of the cane toad (*Bufo marinus*) invasion front in tropical Australia. *Austral Ecology* 32: 169–176.
https://doi.org/10.1111/j.1442-9993.2007.01664.x

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

Reepmeyer, C., S. O'Connor, T. Maloney, and S. Kealy. 2016. Late Pleistocene/early Holocene maritime interaction in Southeastern Indonesia–Timor Leste. *Journal of Archaeological Science* 76: 21–30.
https://doi.org/10.1016/j.jas.2016.10.007

Reutter, B. A., J. Hausser, and P. Vogel. 1999. Discriminant analysis of skull morphometric characters in *Apodemus sylvaticus, A. flavicollis,* and *A. alpicola* (Mammalia; Rodentia) from the Alps. *Acta Theriologica* 44: 299–308.
https://doi.org/10.4098/AT.arch.99-28

R Core Development Team. 2019. *R: a Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Robins, J. H., M. Hingston, E. Matisoo-Smith, and H. A. Ross. 2007. Identifying *Rattus* species using mitochondrial DNA. *Molecular Ecology Notes* 7: 717–729.
https://doi.org/10.1111/j.1471-8286.2007.01752.x

Rozen, S., and H. J. Skaletsky. 2000. Primer3 on the WWW for general users and for biologist programmers. *Methods in Molecular Biology* 132: 365–386.
https://doi.org/10.1385/1-59259-192-2:365

Rousset, F. 2008. genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources* 8: 103–106.
https://doi.org/10.1111/j.1471-8286.2007.01931.x

Schwarz, E., and H. Schwarz. 1967. A monograph of the *Rattus rattus* group. *Annals Escuela Nacional Ciencias Biologicas* 14: 79–178.

St Pierre, E. J. 2011. *U-series Dating and Geochemical Analysis of Speleothems: Developing a Robust Chronological Tool for Cave Deposits and Assessing Late Holocene Human-Environment Interactions in Western Flores, Indonesia*. Unpublished Ph.D. thesis. University of Queensland, Brisbane.

Thomson, V., K. P. Aplin, A. Cooper, S. Hisheh, H. Suzuki, I. Maryanto, G. Yap, and S. C. Donnellan. 2014. Molecular genetic evidence for the place of origin of the Pacific Rat, *Rattus exulans*. *PLoS ONE* 9(3): e91356.
https://doi.org/10.1371/journal.pone.0091356

Thomson, V., A. Wiewel, A. Chinen, I. Maryanto, M. H. Sinaga, R. How, K. Aplin, and H. Suzuki. 2018. A perspective for resolving the systematics of *Rattus*, the vertebrates with the most influence on human welfare. *Zootaxa* 4459(3): 431–452.
https://doi.org/10.11646/zootaxa.4459.3.2

van der Geer, A. A. E. 2018. Changing invaders: trends of gigantism in insular introduced rats. *Environmental Conservation* 45: 203–211.
https://doi.org/10.1017/S0376892918000085

Veatch, E. G., M. W. Tocheri, T. Sutikna, K. McGrath, E. W. Saptomo, Jatmiko, and K. M. Helgen. 2019. Temporal shifts in the distribution of murine rodent body size classes at Liang Bua (Flores, Indonesia) reveal new insights into the paleoecology of *Homo floresiensis* and associated fauna. *Journal of Human Evolution* 130: 45–60.
https://doi.org/10.1016/j.jhevol.2019.02.002

West, K., C. Collins, O. Kardailsky, J. Kahn, T. L. Hunt, D. V. Burley, and E. Matisoo-Smith. 2017. The Pacific rat race to Easter Island: tracking the prehistoric dispersal of *Rattus exulans* using ancient mitochondrial genomes. *Frontiers in Ecology and Evolution* 5: 52.
https://doi.org/10.3389/fevo.2017.00052

Yom-Tov, Y., S. Yom-Tov, and H. Moller. 1999. Competition, coexistence and adaptation among rodent invaders of Pacific and New Zealand islands. *Journal of Biogeography* 26: 947–958.
https://doi.org/10.1046/j.1365-2699.1999.00338.x