Molecular Phylogenetics and Evolution 116 (2017) 227–238

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Tracing the phylogeographic history of Southeast Asian long-tailed macaques through mitogenomes of museum specimens

Lu Yao[a,b,*], Hongjie Li[c,d], Robert D. Martin[a,b,e], Corrie S. Moreau[a,b], Ripan S. Malhi[c,d]

[a] Committee on Evolutionary Biology, University of Chicago, 1025 E. 57th St., Culver Hall 402, Chicago, IL 60637, USA
[b] Integrative Research Center, The Field Museum of Natural History, 1400 S. Lake Shore Dr., Chicago, IL 60605, USA
[c] Department of Anthropology, University of Illinois at Urbana-Champaign, 607 S. Mathews Ave., 109 Davenport Hall, Urbana, IL 61801, USA
[d] Carl R. Woese Institute for Genomic Biology, University of Illinois, 1206 W Gregory Dr., Urbana, IL 61820, USA
[e] Institute of Evolutionary Medicine, University of Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland

ARTICLE INFO

Keywords:
Southeast Asia
High-throughput sequencing
Phylogeography
Intraspecific phylogeny
Macaca fascicularis
Museum specimens

ABSTRACT

The biogeographical history of Southeast Asia is complicated due to the continuous emergences and disappearances of land bridges throughout the Pleistocene. Here, we use long-tailed macaques (*Macaca fascicularis*), which are widely distributed throughout the mainland and islands of Southeast Asia, as a model for better understanding the biogeographical patterns of diversification in this geographically complex region. A reliable intraspecific phylogeny including individuals from localities on oceanic islands, continental islands, and the mainland is needed to trace relatedness along with the pattern and timing of colonization in this region. We used high-throughput sequencing techniques to sequence mitochondrial genomes (mitogenomes) from 95 Southeast Asian *M. fascicularis* specimens housed at natural history museums around the world. To achieve a comprehensive picture, we more than tripled the mitogenome sample size for *M. fascicularis* from previous studies, and for the first time included documented samples from the Philippines and several small Indonesian islands. Confirming the result from a previous, recent intraspecific phylogeny for *M. fascicularis*, the newly reconstructed phylogeny of 135 specimens divides the samples into two major clades: Clade A includes haplotypes from the mainland and some from northern Sumatra, while Clade B includes all insular haplotypes along with lineages from southern Sumatra. This study resolves a previous disparity by revealing a disjunction in the origin of Sumatran macaques, with separate lineages originating within the two major clades, suggesting that at least two major migrations to Sumatra occurred. However, our dated phylogeny reveals that the two major clades split ~1.88 Ma, which is earlier than in previously published phylogenies. Our new data reveal that most Philippine macaque lineages diverged from the Borneo stock within the last ~0.06–0.43 Ma. Finally, our study provides insight into successful sequencing of DNA across museums and shotgun sequencing of DNA specimens as a method to sequence the mitogenome.

1. Introduction

Southeast Asia is made up of a region of the Asian mainland along with thousands of islands varying in size (Fig. 1). Faunas differ fundamentally between the Sunda islands (Sundaland)—the biogeographical region encompassing the continental shelf that was exposed as a continuous landmass during Pleistocene glacial periods—and oceanic islands, those that have never been connected to the mainland (Wallace, 1863). These two major biogeographical regions are separated by what is called the Wallace Line (Fig. 1a). At the southern end, this line separates Bali and Lombok islands at the Strait of Lombok, which is only ~24 km wide. In the north, the separation occurs at the 129–370 km wide Makassar Strait between Borneo and Sulawesi (formerly Celebes) and extends east into the ~201 km wide and ~1500–2500 m deep strait in the Pacific Ocean between Mindanao and the Sanghir Islands (Bergman et al., 1996; George, 1981; Wallace, 1863). The islands to the east of the Wallace Line make up Wallacea. Huxley (1868) corroborated this division but drew the line directly north so that the Philippines, except for Palawan and its associated islands, lie to the east of the Wallace Line.

The colonization patterns and timing in this region are influenced by continuous geographical changes, especially throughout the Pleistocene (Heaney, 1986; Steppan et al., 2003; Jansa et al., 2006; Outlaw and Voelker, 2008; Delson, 1980; Fooden, 2006). Multiple

* Corresponding author at: American Museum of Natural History, Department of Mammalogy, Central Park West and 79th St., New York, NY 10024, USA.
E-mail address: lyao@amnh.org (L. Yao).

http://dx.doi.org/10.1016/j.ympev.2017.08.006
Received 22 November 2016; Received in revised form 23 May 2017; Accepted 13 August 2017
Available online 31 August 2017
1055-7903/

glaciation periods allowed for the Sunda islands to be connected to the current mainland, creating land bridges on which organisms were able to migrate to the various islands. Within some primate groups, there is a unique split within or between species on the continental island of Sumatra. Specifically, a southernmost Sumatran orangutan population of the species *Pongo abelii* is in fact more closely related to the Bornean orangutan, *Pongo pygmaeus*, than it is to the northern populations of *P. abelii* (Nater et al., 2011, 2015). This divergence is also seen in gibbons, with the white-handed gibbon (*Hylobates lar*) occurring to the north and the mountain agile gibbon (*Hylobates agilis*) occurring to the South on Sumatra (Whittaker et al., 2007; Thinh et al., 2010). But there do not appear to be other studies that have identified splits like these in other taxa on Sumatra.

It is unlikely that any major land bridges formed between the Philippines and Sundaland, aside from a possible connection between Borneo and Palawan (Heaney, 1985). So humans are thought to have influenced the colonization of oceanic islands by bringing non-volant organisms to these islands in the very recent past (Heaney et al., 2016). However, if specific species were able to migrate to the Philippine islands without human aid, it is possible they would be able to easily colonize this oceanic region since several of the islands within the Philippines were likely connected via land bridges during glaciation periods (Heaney, 1985). In order to shed light on the patterns and timing of colonization throughout both Sunda and oceanic islands in Southeast Asia, it is necessary to study an organism that is widespread throughout the region. There appears to be a disjunction in faunal occurrences on either side of Huxley's line, but *M. fascicularis* is an exception as one of the few species of non-volant mammals that occurs widely on the mainland, Sundaland, and on oceanic islands, making this species an excellent organism for studying biogeographical diversification throughout this entire region (Fig. 1).

Macaca species belong to one of the most widely distributed primate genera, occurring on two continents and multiple islands. Fossil evidence indicates that this highly successful genus originated in Africa around 7 million years ago (Ma) (Delson, 1980), after which macaques expanded into Asia approximately 5.5 Ma (Alba et al., 2014; Delson, 2000). Twenty-two species subdivided into seven species groups are currently recognized based on distribution, morphology, behavior and genetics (Li et al., 2009; Tosi et al., 2003). Based mostly on morphological data, three of these groups are monospecific: *M. sylvanus*, which is the only extant macaque in northern Africa and southern Europe, *M. fascicularis*, and *M. arctoides*. The remaining four groups are polyspecific, with six species in the Sulawesi group, five in the *M. silenus* group, three in the *M. mulatta* group, and five in the *M. sinica* group (Zinner et al., 2013a). However, classification into species groups has changed over time in tandem with extensive debate. Originally, *M. mulatta*, *M. cyclopis*, *M. fuscata*, and *M. fascicularis* were all included in the *M. fascicularis* species group (Fooden, 1976). Groves (2001) and Zinner et al. (2013a) then combined *M. mulatta*, *M. cyclopis*, and *M. fuscata* in a *M. mulatta* species group. Groves (2001) also added *M. arctoides* to the *M. fascicularis* species group, but Zinner et al. (2013a) subsequently classified *M. arctoides* and *M. fascicularis* as two separate

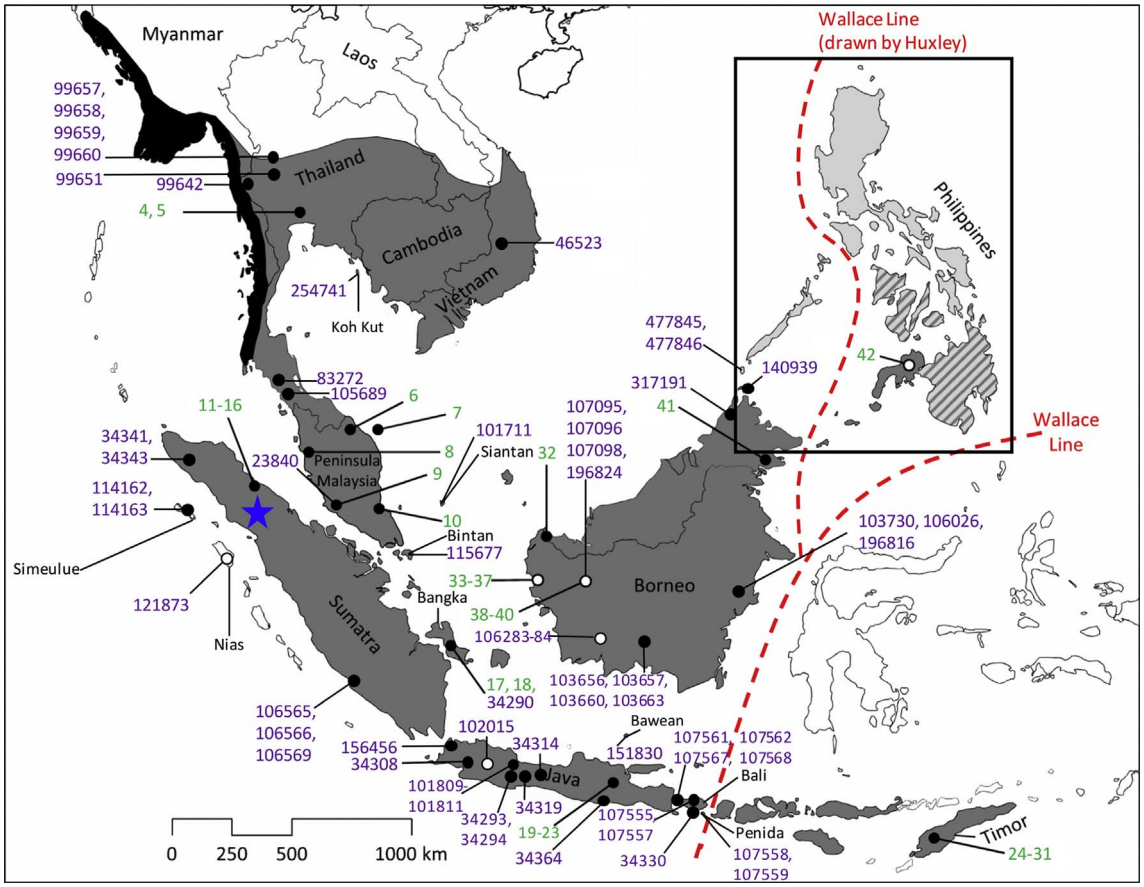

Fig. 1. Maps of Southeast Asia. (a) A map of Southeast Asia derived from Liedigk et al. (2015) marked with specimens from the current study and Liedigk et al.'s (2015) study. The black, dark grey and light grey regions respectively indicate the ranges of *M. f. aureus*, *M. f. fascicularis* and *M. f. philippinensis*. The region colored in light and dark grey lines is the putative area of intergradation of *M. f. fascicularis* and *M. f. philippinensis* based on studies by Fooden (1995, 2006). White circles represent approximate location whereas black circles represent exact location. ID numbers of *M. fascicularis* museum specimens in green correspond to samples sequenced in Liedigk et al. (2015) and those in purple are the newly sequenced samples in this study. Lake Toba is marked with a blue star. (b) A map of the Philippines (Philippines region in rectangular box from (a) magnified), derived from Heaney (1986), marked with specimens from the current study. The circles and colored ID numbers of *M. fascicularis* specimens are labeled according to the description in map (a). All ID numbers correspond to those in Fig. 2 and Supplementary Data 1. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

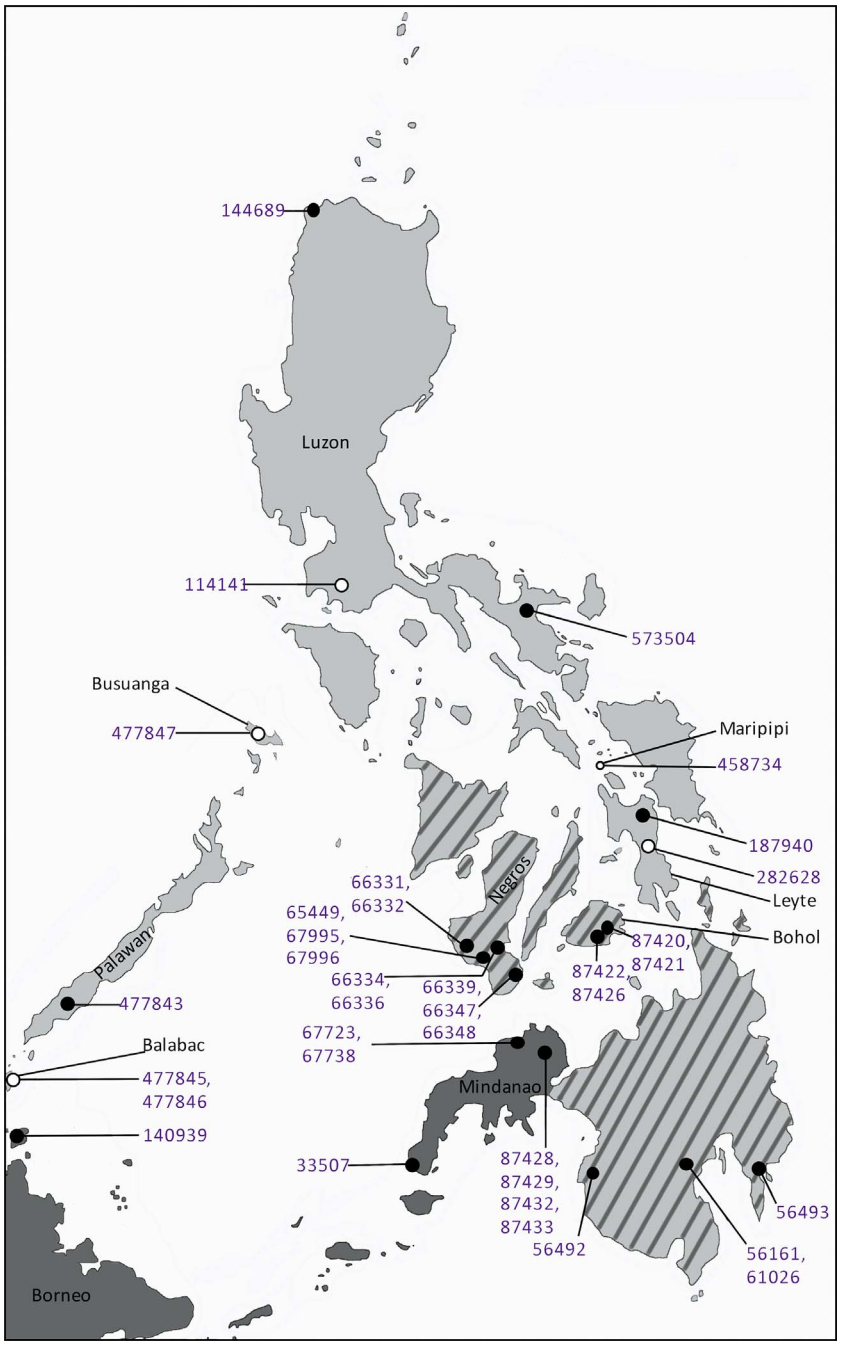

Fig. 1. (*continued*)

monotypic groups.

A study investigating species-level relationships within the genus *Macaca* by detecting the presence and absence pattern of *Alu* elements, the most abundant short interspersed elements (SINEs) in primates, concluded that *M. fascicularis* is the sister group to a clade containing *M. fuscata* and *M. mulatta* (Li et al., 2009). Detailed studies have demonstrated that there have been major hybridization events between *M. fascicularis* and *M. mulatta*, most likely reflecting overlapping geographical distribution on the Indo-Chinese peninsula (Bonhomme et al., 2009; Fooden, 1964, 2000; Kanthaswamy et al., 2009; Stevison and Kohn, 2009; Tosi et al., 2002). In a study comparing the genomes and divergence rates of these two species, Yan et al. (2011) found that approximately 30% of the *M. fascicularis* genome originates from *M. mulatta*. This gene flow appears to be non-maternal in origin as mitochondrial analyses show no signs of such hybridization. *Macaca fascicularis* and *M. nemestrina* have overlapping geographical distributions on the Southeast Asian islands, but no cases of natural hybridization between these two species have been reported.

Ten subspecies have been recognized within the species *M. fascicularis*, but there are few visible differences aside from pelage color, tail length and the shape of the cheek whiskers (Groves, 2001; Ong and Richardson, 2008). Seven of these subspecies are allopatric, with localized distributions on small islands (*M. f. atriceps, M. f. condorensis, M. f. fuscus, M. f. karimondjawae, M. f. lasiae, M. f. tua, M. f. umbrosus*), while the remaining three are widely distributed throughout Southeast Asia (*M. f. fascicularis*, *M. f. philippinensis*, *M. f. aureus*). *Macaca f. aureus* is geographically isolated from *M. f. fascicularis* and *M. f. philippinensis,* but there is a putative area of intergradation of *M. f. fascicularis* and *M. f. philippinensis* on some of the Philippine islands, including southern Negros and Mindanao (Fooden, 1995; see Fig. 1).

It has been inferred that *M. fascicularis* first colonized Sundaland during global cooling in the Pliocene (~5.3–2.6 Ma) (Delson, 1980), and that subpopulations became isolated on those islands once sea levels rose. During the Pleistocene glacial periods, *M. fascicularis* was able

to expand its range even further throughout the continental shelf islands (Delson, 1980; Fooden, 2006), as is evidenced by fossil remains of *M. fascicularis* on Java dated to the early Holocene or middle Pleistocene (Aimi and Aziz, 1985; Fooden, 2006) and the extensive genetic diversity in Sunda shelf populations in comparison to those on the mainland (Smith et al., 2007). Time-calibrated molecular phylogenies indicate a time frame within the last 5 million years for the arrival of both *M. fascicularis* and other primates on the larger Sunda shelf islands (Chan et al., 2010; Chatterjee, 2006; Liedigk et al., 2015; Tosi and Coke, 2007). Tosi and Coke (2007) inferred from a 1.5 kb fragment of mtDNA and two Y-chromosome loci that *M. fascicularis* colonized Sumatra in multiple waves throughout the Pleistocene, starting at about 1.2 Ma and continuing until the most recent glacial maximum. This study analyzed the mtDNA fragment and Y-chromosome loci separately. mtDNA analyses indicated that separation of macaques on the Sunda shelf islands from the mainland occurred approximately 1.2 Ma, and there was later a bifurcation in the Y-chromosome loci at around 0.4 Ma. The separation based on mtDNA is the deepest intraspecific bifurcation within *M. fascicularis*. Moreover, the mtDNA-based phylogeny reconstructed by Tosi and Coke (2007) clustered Sumatran individuals with conspecifics on other islands, whereas the Y-chromosome analyses indicate that the Y-chromosomal loci of Sumatran individuals cluster with those of mainland individuals as well as with other insular individuals. The authors reconciled this difference with the explanation of secondary contact. Overall, they found that *M. fascicularis* individuals from Sumatra are split into some that cluster with other islands and some that cluster with the Southeast Asian mainland in general.

The most recently published intraspecific phylogeny for *M. fascicularis* is based on entire mitogenomes (Liedigk et al., 2015). Using five fossil calibrations and a relaxed molecular clock approach, Liedigk et al. found the divergence between the mainland-Sumatran clade and the clade containing other insular individuals to be ~1.7 Ma. However, their results differ from those of Tosi and Coke's (2007) study in that all Sumatran individuals represented cluster with the mainland samples rather than the insular individuals. Their mitogenome phylogeny again indicates the presence of two separate clades. Clade A includes all mainland and Sumatran lineages in a paraphyletic array, while Clade B includes the lineages from all other islands in monophyletic groups except for their only Philippine individual, which was nested within the Borneo clade.

Even though *M. fascicularis* is widespread throughout most of Southeast Asia, very few studies have examined the timing of colonization of oceanic islands by this species (Fooden, 2006). Additionally, intraspecific relationships of *M. fascicularis* have been resolved only for the mainland, the larger Sunda islands and a single Philippine island (Liedigk et al., 2015; Tosi and Coke, 2007). A genetic study of *M. fascicularis* in the Philippines concluded that those populations had naturally colonized the islands in two major waves separated in time (Smith et al., 2014). On the other hand, Fooden (2006) and Heaney et al. (2016) hypothesized that the lineages on the Philippines east of the Huxley-Wallace Line were introduced by humans approximately 4000–3500 years ago. Because *M. fascicularis* are able to swim well, it is thought that they could have swum between islands that are in close proximity (Gumert and Malaivijitnond, 2012). The single Philippine individual included in the study by Liedigk et al. (2015) apparently diverged from the Borneo lineages approximately 0.21 Ma, considerably earlier than the hypothesized date for human introduction. The exact origin of this Philippine individual was unclear, although it was marked as a specimen from Mindanao on their map (Fig. 1 in Liedigk et al., 2015).

The aim of the present study was to better understand the biogeographical diversification of mammals throughout Southeast Asia by using *M. fascicularis* as a model organism due to its widespread geographical range. Understanding the phylogeographic relationships within this species provides clarification of colonization patterns and timing in a region with a complicated biogeographical history. To achieve this objective, we used tissue samples from museum specimens of *M. fascicularis* from many distinct localities throughout the Southeast Asian region, focusing particularly on the Philippines and small Sunda islands, and sequenced their entire mitochondrial genomes (mitogenomes) using high-throughput sequencing. This more comprehensive intraspecific phylogeny provides a more robust understanding of colonization of this extensive geographical region by long-tailed macaques.

2. Methods

All molecular lab work along with associated computational processing was conducted at two major locations: (1) Malhi ancient DNA lab and the Malhi Molecular Anthropology lab at the University of Illinois at Urbana-Champaign (UIUC), and (2) the Pritzker Lab for Molecular Systematics and Evolution at the Field Museum of Natural History (FMNH). To minimize contamination, at each extraction and amplification setup step a clean-room facility (the Malhi ancient DNA lab), contamination prevention protocols, and negative experimental controls were used (as in Lindo et al., 2016). No animals were sacrificed for this study.

Mitochondrial DNA (mtDNA) is widely used in population genetic studies because its sequences evolve relatively rapidly and are inherited maternally, such that they lack recombination (Brown et al., 1979). Because evolution on islands may occur comparatively quickly (Evans et al., 2012), mtDNA is an excellent tool for studying diversity within populations of Southeast Asian mammals. Specifically, we examined the mitogenome of *M. fascicularis* (approximately 16,500 bp) using next-generation sequencing, because the mitogenome can provide finer phylogenetic resolution and precision compared to traditional localized mtDNA markers. Using different regions of the mitogenome can yield incongruent results concerning divergence dates, taxonomy and phylogeography (Pacheco et al., 2011; Rohland et al., 2007). Additionally, there are numerous copies of the mitogenome in each cell whereas there are only two copies of the nuclear genome in diploid organisms, making the mitogenome easier to sequence, especially with low-quality samples such as in ancient DNA (Briggs et al., 2009; Guschanski et al., 2013; Higuchi et al., 1984; Krause et al., 2010; Mason et al., 2011; Pääbo, 1989; Pääbo et al., 1989; Rowe et al., 2011).

2.1. Sample collection

Recent tissue samples from Southeast Asian *M. fascicularis* are difficult to obtain, especially from museum collections, because the most recent specimens collected from the region (late 1990s) were not prepared for DNA extraction, and current regulations now restrict tissue collection and importation from live primates. Samples for DNA sequencing in this project were therefore collected from fragments of dried tissue of 50- to 150-year-old wild *M. fascicularis* specimens at the Field Museum of Natural History (FMNH) in Chicago, American Museum of Natural History (AMNH) in New York, Smithsonian Institution National Museum of Natural History (NMNH) in Washington, and the Naturalis Biodiversity Center (RMNH) in Leiden, Netherlands. Skins and toe pads were not sampled because some museums forbid this type of sampling from all specimens within a species, and other scientists study these anatomical parts, whereas the dried tissues left on skeletal remains are normally not used for any other analyses. Using a scalpel and tweezers, which were sanitized between specimens with heat or bleach, dried tissue fragments comparable in size to 1–2 grains of rice were carefully excised either from cranial bone surfaces or within the braincase of both male and female skulls of museum specimens. Efforts were made to collect tissue samples from as many localities as possible. Tissue samples representing 196 specimens were obtained from the mainland and from 22 Southeast Asian islands (see Fig. 1 and Supplementary Data 1 for geographical localities of successfully sequenced specimens in this study).

2.2. DNA extraction

DNA was extracted from the tissue samples following the ancient DNA extraction protocol established in the Malhi Lab (Cui et al., 2013). To eliminate contamination problems, all extraction work was performed in the sterile ancient DNA lab at UIUC. Dried tissue samples were first digested overnight in a rotating 37 °C incubator in 4 mL 0.5 M EDTA, 150 µl of 33.3 mg/ml proteinase K and 300 µl of 10% N-lauryl sarcosine. All DNA extractions were accompanied by negative controls to permit detection of contamination during the extraction process. Once the samples were digested, DNA was extracted using silica columns and reagents from a Qiagen extraction kit following the manufacturer's protocol and eluted to 100 µl per sample.

Because DNA from dried tissue is fragmented and often degraded, it is important to confirm that usable DNA is present. Accordingly, we designed primers for a 200 bp section of *M. fascicularis* cytochrome *b* (cyt*b*) based on previously published mitogenomes (Liedigk et al., 2015). These primers (MFAS_F: 5′-TACGCAAATCCAACCCAATC-3′; MFAS_R: 5′-GGTGATGTGTGCAATTGAGG-3′) were successfully tested on *M. fascicularis* tissues excised from two frozen cadavers in Dr. Callum Ross's biomechanics laboratory at the University of Chicago and on three ancient DNA samples. We set up each sample of extracted DNA with negative controls in a PCR reaction using these primers. These samples were then taken in an airtight container to the Malhi Molecular Anthropology laboratory two blocks away where a positive control was added. The samples were placed in an Eppendorf Mastercycler (thermal cycler) for Polymerase Chain Reaction (PCR) amplification using an optimized 52 °C annealing temperature and 35 cycles.

All PCR-amplified samples along with negative and positive controls were run on agarose gels in the Pritzker Lab to determine the samples with the best preserved DNA. We then proceeded with the preparation of genomic libraries for those specimens.

2.3. Genomic libraries and high-throughput sequencing

Genomic libraries with Illumina platform-specific oligonucleotide adapters unique to each library were created in the ancient DNA lab using 50 µl of extracted DNA per sample and the NEBNext Ultra DNA Library Prep Kit for Illumina following the TruSeq DNA Sample Preparation V2 protocol by Illumina. Due to the particular nature of DNA extracted from museum specimens, certain modifications were made to the protocol. First, the DNA extract was not sheared because the DNA is already highly fragmented. The concentration of DNA in each extract was expected to be low, so the adapters were diluted 1:20 (Cui et al., 2013). Adapter dimers often form during ligation, so multiple AMPure Bead XP clean ups were conducted.

PCR setup of the genomic libraries using unique NEBNext Multiplex Dual Index primers was also conducted in the Malhi ancient DNA laboratory. Dual index primers were used to avoid cross-contamination when multiplexing samples on a single HiSeq lane (Kircher et al., 2012). The samples were transferred to thermal cyclers in an airtight container in the Malhi Molecular Anthropology lab for amplification. NEBNext High Fidelity 2X PCR Master Mix was used to amplify the libraries because its proof-reading properties limit nucleotide misincorporations resulting from cytosine deamination (Ginolhac et al., 2011). The amplified genomic libraries were cleaned using the Qiagen MinElute Purification Kit.

The libraries were then assessed for fragment size and quantification using the Agilent 2100 Bioanalyzer or the AATI Fragment Analyzer. Nucleic acid concentrations were determined using a Qubit 2.0 Fluorometer. Finally, the samples were pooled and shotgun sequenced at the Keck Biotech Center at UIUC, using an ILLUMINA HiSeq2500, which generated lanes of 100-bp single-end reads for 24–30 samples per lane.

2.4. Alignment and assessment

The Trimmomatic program was used to trim DNA sequences and remove adaptors along with the sequences that entered the sample after the clean room preparation. This step reduces false variant discovery by ignoring reads that are below standard quality due to DNA damage (Kircher et al., 2012). BOWTIE 2 (Langmead and Salzberg, 2012) was used to assemble each sample against a *M. fascicularis* reference mitogenome (GenBank ID: KJ567052.1; Liedigk et al., 2015). The program SAMtools was utilized for sorting, indexing, and quantifying contamination by examining informative sites (Malström et al., 2007) and removing potential duplicate reads that could result from PCR amplification.

2.5. Phylogenetic reconstruction

To reconstruct the phylogenetic relationships of *M. fascicularis* we included mitogenome sequences for 95 samples successfully sequenced as part of this current study (Supplementary Data 1), 40 *M. fascicularis* mitogenome sequences from Liedigk et al. (2015), and 17 mtDNA genome sequences from other macaques (10 individuals from six species) and non-macaque primate taxa (seven individuals from different genera) included in Liedigk et al. (2015) as outgroups to expand the phylogenetic perspective.

All sequences were aligned with MAFFT version 7 (Katoh, 2013) and corrected by hand in Mesquite (Maddison and Maddison, 2011). We then manually removed poorly aligned positions and indels along with missing data from coding regions. For maximum likelihood and Bayesian analyses, we evaluated two different partition schemes of the mitogenome: (1) no partition and (2) 66 partitions, which included the 2 rRNA partitions, 22 tRNA partitions, one partition per codon for the D-loop (three partitions), and one partition per codon for each of the 13 different protein-coding genes (39 partitions). The nucleotide substitution model for the two partition schemes was selected from 56 models using the corrected Akaike Information Criterion (AICc) implemented in PartitionFinder 2.1.1 (Lanfear et al., 2012).

To perform the maximum likelihood and Bayesian reconstructions, we used RAxML 8.2.4 (Stamatakis, 2006), and MrBayes 3.2.2 (Ronquist and Huelsenbeck, 2003), respectively, through the CIPRES Science Gateway V3.3 (Miller et al., 2010). Both RAxML and MrBayes were run using the GTR + I + G model based on results from PartitionFinder (Supplementary Data 2). Maximum likelihood calculations were run with 100 bootstrapping replications in RAxML. The parameters for MrBayes included four Markov Chain Monte Carlo (MCMC) runs. All analyses were run for 10 million generations with a parameter sampling frequency of 1000 generations, with the first 10% of samples discarded as burn-in.

In order to estimate divergence dates, we implemented BEAST 1.8.3 (Drummond and Rambaut, 2007; Suchard and Rambaut, 2009) with the Bayesian MCMC method and relaxed molecular clock model (Drummond et al., 2006) using CIPRES Science Gateway V3.3 (Miller et al., 2010). We assumed a Birth-Death Process prior for branching rates, as was previously done in Liedigk et al. (2015). We used the same five fossil calibration points as in Liedigk et al. (2015) with a lognormal distribution prior for all nodes: (1) *Homo* – *Pan* divergence at 6.5 Ma with 95% CI of 0.5 Ma (Brunet et al., 2005; Lebatard et al., 2008; Vignaud et al., 2002), (2) *Pongo* – *Homo* + *Pan* clade divergence at 14 Ma with 95% CI of 1.0 Ma (Kelley, 2002), (3) *Theropithecus* – *Papio* divergence at 5 Ma with 95% CI of 1.5 Ma (Delson, 2000; Leakey, 1993), (4) African – Asian macaques at 5.5 Ma with 95% CI of 1.0 Ma (Delson, 2000; Alba et al., 2014), and (5) hominoids – cercopithecoids at 27.5 Ma with 95% CI of 3.5 Ma (Stevens et al., 2013 and Zalmout et al., 2010; but see also Pozzi et al., 2011). Hard lower boundary constraints were applied to all fossil calibrations as the true divergence can only be older than the fossil, not younger. A total of four replicates were run for 700 million generations for the non-partitioned dataset

and 600 million generations for the partitioned dataset, and the tree and parameter sampling was performed every 5000 generations. We used Tracer to check that a 10% burn-in is sufficient and all ESS values suggested convergence. TreeAnnotator 1.8.3 was used to reconstruct a consensus topology based on the distribution of trees, and we visualized the phylogeny with FigTree 1.4.2 (Rambaut). We also conducted a pairwise distance analysis in PAUP (Swofford, 2002) to check if any specimens were driving an increase in diversity.

3. Results

From a total of 196 tissue samples collected, we successfully extracted DNA from 151 (77%), of which the mitogenomes of 95 *M. fascicularis* individuals were successfully sequenced by shotgun sequencing using the Illumina HiSeq2500 (see Supplementary Data 1 for list of specimens sequenced here and their Genbank accession numbers). All raw sequences are available on Genbank. The mitogenomes of 56 samples for which we had built libraries were not fully sequenced and had relatively low-quality scores and were excluded from the analyses. The bioanalyzer or AATI results for the overall set of samples were not indicative of how well the libraries would perform during shotgun sequencing. On seven HiSeq 2500 lanes, we obtained an average of 674,364 (49,852–2,064,668) total reads. The average percentage of mitogenome reads compared to the total reads is 0.14% (0.01–0.9%), with an average 40.95 × (3.28 − 125.03) coverage for the mitogenomes that were successfully sequenced per sample. All mitogenomes that we generated and included in the following analyses ranged between 15,185 and 16,690 bp. It is important to note that there was a range of extraction and sequencing success for the museums where they were collected (Table 1).

All steps that we took to check for contamination gave negative results. At the extraction stage, all negative controls were found to be clean, and the positive controls using DNA extracted from the two fresh cadavers of *M. fascicularis* indicated that the extraction process was successful when amplifying the 200 bp of Cyt*b*. After sequencing, the BLAST results for each sequence revealed that the sequences of interest matched those of *M. fascicularis* mitogenomes.

In total, we aligned 95 newly generated *M. fascicularis* mitogenomes (Supplementary Data 1), 40 *M. fascicularis* mitogenomes previously sequenced by Liedigk et al. (2015), and 17 mitogenomes of outgroup taxa. The latter two sets of mitogenomes were downloaded from Genbank with the accession numbers provided in Liedigk et al. (2015). The final alignment had a total length of 16,690 bp (Supplementary Data 3). Some portions of the mitogenome from five specimens, specifically from Bali (AMNH 107568), Simeulue (NMNH 114163), Bawean (NMNH 151830), Banggi (FMNH 140939), and Java (RMNH 34314), were incomplete. However, the missing base pairs were not from the same regions. We do not believe that missing short sections of the mitogenomes in those specimens severely affected the results at the intraspecific level because only 1505 bp were missing at most from a single specimen, and those specimens cluster within the phylogeny as would be expected (see Fig. 2). One specimen from Thailand (NMNH 251661) that was sequenced, which is not included in the 95 specimens, was the only specimen found to be an outgroup to all *M. fascicularis* samples. Although no other specimens from the same locality were sequenced to test whether or not this one specimen is representative of its source population, the specimens from nearby localities do not cluster with this single specimen from Nakhon Si Thammarat in Thailand. Therefore, this specimen was removed from the alignment and further phylogenetic analyses.

Nearly identical phylogenies with strongly supported nodes (ML bootstrap values: > 95%, Bayesian posterior probabilities: 1.0) were reconstructed based on maximum-likelihood and Bayesian analyses (Supplementary Data 4 and 5, respectively). The newly reconstructed phylogeny included samples from regions with known localities that had not previously been sampled at the mitogenome level. They include DNA of *M. fascicularis* from small Sunda islands such as Bali, Bintan, Banggi, Penida, Siantan, Bawean, Nias, Simeulue, and Koh Kut islands and from the Philippine islands, which include Mindanao, Negros, Bohol, Maripipi, Luzon, Leyte, Palawan, Busuanga, and Balabac islands. The phylogeny splits the specimens into two major clades, with Clade A from Liedigk et al. (2015) containing all mainland specimens, northern Sumatran specimens, specimens from Bintan, Siantan and Koh Kut islands, and a single specimen from Mindanao. Clade B contains all other insular specimens including those from southern Sumatra.

Based on the analyses in this study, the divergence dates for most divergences are earlier than had been previously estimated but overlap in their confidence intervals (Fig. 2; Table 2; Supplementary Data 6 and 7). The following are the divergence dates for the phylogeny reconstructed from the partitioned dataset, which are very similar to those in the phylogeny reconstructed from the non-partitioned dataset (see Table 2 and Supplementary Data 6 and 7). At the root, Hominidae and Cercopithecidae separated 31.16 Ma (95% credibility interval [CI]: 27.95–36.57). Within the hominids, *Homo* and *Pan* split 6.98 Ma (6.68–7.40), and *Pongo* diverged from *Homo* + *Pan* 14.94 Ma (14.29–15.98). Within Cercopithecidae, *Colobus* diverged first 22.52 Ma (19.18–26.45) followed by the divergence of *Chlorocebus* 14.70 Ma (12.74–17.01). Papionini split from *Macaca* 12.50 Ma (10.81–14.40), and within Papionini *Theropithecus* and *Papio* diverged 5.64 Ma (5.23–6.21). Within the macaques, the African and Asian macaques diverged 6.83 Ma (6.02–7.84). *M. silenus* and *M. tonkeana* diverged next from the other Asian macaques 6.15 Ma (5.37–7.08). These two species split from each other 4.18 Ma (3.41–5.01). *M. thibetana* split from the group including *M. fascicularis, M. mulatta* and *M. arctoides* 4.65 Ma (3.98–5.39). *M. fascicularis* then diverged from *M. mulatta* and *M. arctoides* 3.85 Ma (3.28–4.49) and these latter two species split 3.42 Ma (2.86–4.04).

Within *M. fascicularis,* Clades A and B diverged early at 1.88 Ma (1.60–2.20). Clade A's splitting events began approximately 1.03 Ma (0.88–1.20). Within this clade, the individuals are not grouped into monophyletic clades by geographic localities. The northern Sumatran specimens that we sequenced form a cluster, which also contains lineages from mainland Malaysia, Thailand, and Bintan and Siantan islands, diverging from mainland populations between 0.25–0.43 Ma

Table 1
Extraction and sequencing success of specimens. Number of *M. fascicularis* samples collected and percentage of successful specimens sequenced for samples from four museum collections. FMNH = The Field Museum of Natural History in Chicago, IL; AMNH = American Museum of Natural History in New York, NY; NMNH = Smithsonian Institution National Museum of Natural History in Washington, DC; RMNH = Naturalis Biodiversity Center in Leiden, Netherlands. % sequenced represents the % of successfully sequenced samples of the successfully extracted samples.

Museum	# Sampled	# Successful extractions	% Extracted	# Successfully sequenced	% Sequenced
FMNH	71	55	77	34	62
AMNH	51	34	67	28	82
NMNH	54	44	81	22	50
RMNH	20	19	95	11	58
Total	196	152	78	95	63

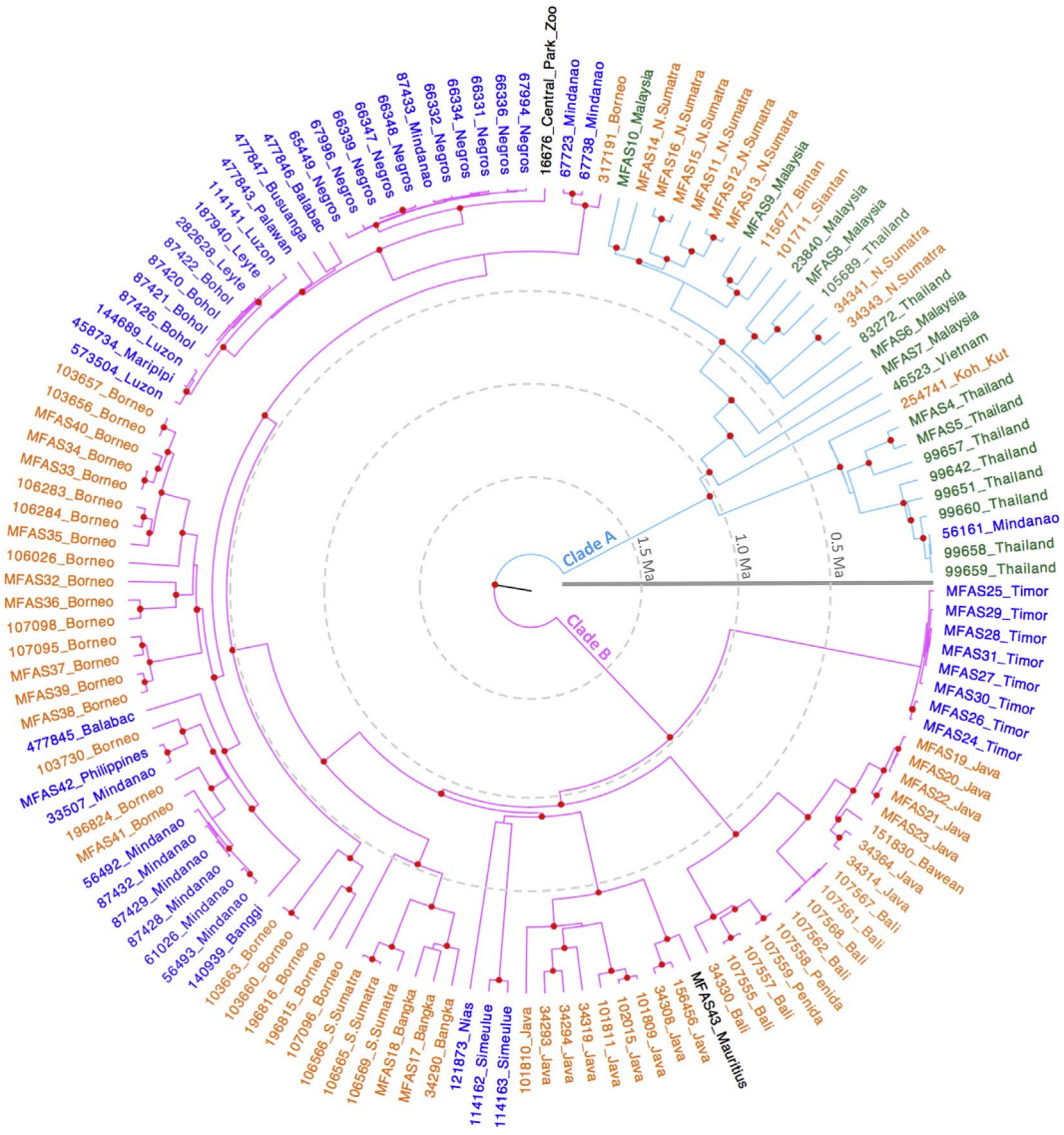

Fig. 2. Phylogeny of *Macaca fascicularis*. Newly reconstructed phylogeny of 135 M. *fascicularis* museum specimens with divergence dates based on partitioned dataset. Samples starting with MFAS are from Liedigk et al. (2015) while samples beginning with museum specimen numbers were newly sequenced for this study. Green = mainland, orange = Sundaland, blue = oceanic islands. All significant nodes are marked with a red circle. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

(0.20–0.51). There are four other insular specimens within Clade A, and those include specimens from Koh Kut, Siantan, Bintan, and Mindanao islands. The divergence of the specimen in Mindanao from Thailand specimens is the most recent of the four, at 7.6 kya (1.4–16.5 kya).

Liedigk et al. (2015) had found Clade B to contain distinct monophyletic lineages based on geographic regions, but the results in this study reveal that this is not the case, especially for specimens from Borneo, the Philippines, and Java and its surrounding islands. The Timor specimens that Liedigk et al. (2015) sequenced diverged as a single clade within Clade B 1.03 Ma (0.88–1.19). The specimens from Java and its surrounding islands diverged as two major groups. The first group, which contains solely specimens from Java and its surrounding islands, diverged 0.96 Ma (0.82–1.10). The second group, which also includes specimens from Nias, Simeulue and Mauritius islands, then diverged 0.93 Ma (0.80–1.07). The Mauritius lineage split off 0.14 Ma (0.10–0.19), while the monophyletic group of specimens from Nias and Simeulue islands diverged from the Javan specimens earlier at 0.91 Ma (0.78–1.05). The split of the lineages from these two islands to the west of Sumatra occurred 0.87 Ma (0.74–1.01). After the paraphyletic Javan clade, a group containing specimens from Borneo, Bangka and southern Sumatra diverged within Clade B 0.69 Ma (0.59–0.80). The Bangka and southern Sumatra lineages diverged from a western Borneo lineage 0.55 Ma (0.45–0.65), and the former lineages split 0.37 Ma (0.29–0.44). The clade containing specimens from Borneo and the Philippine islands began splitting 0.43 Ma (0.37–0.50). The Philippine specimens do not form a monophyletic clade. One group containing lineages from Banggi, Balabac and Mindanao islands diverged from other Bornean specimens 0.38 Ma (0.32–0.44). Additionally, there are Borneo specimens nested within this particular group. The remaining Philippine specimens form a monophyletic group with the exception of a single specimen from northern Borneo, which diverged from two Mindanao specimens 0.07 Ma (0.04–0.10). The lineages from Negros Island diverged from other Philippine specimens 0.15 Ma (0.12–0.18). Although these divergence dates are earlier than previously noted in other studies of *M. fascicularis*, the results of the pairwise distance analysis in PAUP (Swofford, 2002) show that no specimens stood out in the analysis (Supplementary Data 8).

Table 2
Divergence dates. Comparison of major divergence dates for *M. fascicularis* inferred in Liedigk et al. (2015) study and in the present study. Divergence dates for partitioned and non-partitioned dataset reconstructions are provided.

Clades	Liedigk et al. (Ma)	Current study (Ma) with partitioning	Current study (Ma) without partitioning
Hominidae – Cercopithecidae	28.60 (25.31–31.78)	31.16 (27.95–36.57)	31.13 (27.99–36.35)
Pongo – *Homo* + *Pan*	13.82 (12.68–14.86)	14.94 (14.29–15.98)	14.92 (14.27–15.96)
Homo – *Pan*	6.32 (5.73–6.89)	6.98 (6.68–7.40)	6.98 (6.69–7.40)
Colobus – other Cercopithecidae	19.89 (16.17–23.87)	22.52 (19.18–26.45)	22.40 (18.96–26.45)
Chlorocebus – Papionini + *Macaca*	12.81 (10.59–15.22)	14.70 (12.74–17.01)	14.70 (12.53–16.92)
Papionini – *Macaca*	10.90 (8.92–12.90)	12.50 (10.81–14.40)	12.51 (10.73–14.38)
Theropithecus – *Papio*	4.77 (3.87–5.72)	5.64 (5.23–6.21)	5.65 (5.23–6.24)
M. sylvanus – Asian macaques	6.10 (5.23–6.92)	6.83 (6.02–7.84)	6.83 (6.02–7.80)
M. silenus + *M. tonkeana* – other Asian *Macaca*	5.49 (4.69–6.34)	6.15 (5.37–7.08)	6.15 (5.38–7.04)
M. silenus – *M. tonkeana*	3.70 (2.80–4.54)	4.18 (3.41–5.01)	4.19 (3.41–5.00)
M. thibetana – *M. fascicularis* + *M. mulatta* + *M. arctoides*	4.16 (3.47–4.85)	4.65 (3.98–5.39)	4.66 (4.01–5.38)
M. fascicularis – *M. mulatta* + *M. arctoides*	3.42 (2.83–4.01)	3.85 (3.28–4.49)	3.86 (3.30–4.47)
M. mulatta – *M. arctoides*	3.02 (2.42–3.60)	3.42 (2.86–4.04)	3.43 (2.88–4.00)
Clade A – Clade B	1.70 (1.36–2.04)	1.88 (1.60–2.20)	1.88 (1.60–2.19)
Clade A splitting events	0.96 (0.78–1.16)	1.03 (0.88–1.20)	1.03 (0.88–1.19)
N. Sumatran specimens – mainland specimens	–	0.25–0.43 (0.20–0.51)	0.25–0.43 (0.20–0.50)
Clade B splitting events (Timor specimens – rest of Clade B)	0.93 (0.74–1.12)	1.03 (0.88–1.19)	1.02 (0.88–1.82)
Javan and surrounding insular specimens – rest of Clade B (excluding Timor specimens)	0.87 (0.70–1.05)	0.93–0.96 (0.80–1.10)	0.93–0.95 (0.80–1.10)
Nias and Simeulue specimens – Javan specimens	–	0.91 (0.78–1.05)	0.90 (0.78–1.04)
S. Sumatran + Bangka specimens – Borneo specimens	–	0.55 (0.45–0.65)	0.54 (0.45–0.64)
S. Sumatran specimens – Bangka specimens	–	0.37 (0.29–0.44)	0.36 (0.30–0.44)
Philippine specimens – Borneo specimens	0.21 (0.15–0.28)	0.06–0.43 (0.04–0.50)	0.06–0.43 (0.04–0.50)

4. Discussion

DNA extracted from museum specimens is usually highly degraded (Burrell et al., 2015; Liedigk et al., 2015; Mason et al., 2011), but with the development of next-generation sequencing techniques it has become faster and less costly to sequence such fragmented DNA. Although many studies use DNA-capture techniques to target the region of interest, such as the mitogenome, we have found that shotgun sequencing the genomes will yield high coverage of entire mitogenomes if the genetic libraries are of high quality. An additional advantage of shotgun sequencing is that it is possible to sequence the nuclear genomes at low coverage (approximately 1x if multiplexing samples in this particular study), which can be used for further studies. But, as past studies have emphasized, there are only two copies of the nuclear genome in a cell in comparison to the tens to thousands of copies of mitogenomes in each cell, so it is inherently more difficult to achieve coverage of the nuclear genome than the mitogenome when sequencing ancient DNA (Guschanski et al., 2013; Hagelberg and Clegg, 1991; Mason et al., 2011; Rowe et al., 2011). As we have found, next-generation sequencing does not achieve high-quality and/or high-precision results for every specimen. Few studies have looked at extraction and sequencing success of specimens across museums and collectors (Bailey et al., 2015). Because we collected tissues from the same species in the same region from four different natural history museums, this study sheds some light on variation in the success of sequencing museum specimens across various collections (Table 1). The highest success for extractions of high quality DNA (95%) was for specimens from the collections at RMNH, and the lowest extraction of high quality DNA (67%) was for specimens housed at AMNH. Sequencing success of the entire mitogenome differed from that for extraction, with the greatest success (82%) in AMNH specimens and the lowest success (50%) for NMNH specimens. The differences in successful extraction and sequencing among various museums may be due to the effects of various chemicals used for cleaning the skeletal materials on the preservation of dried tissues. Unfortunately, how the specimens were cleaned and prepared when they were first collected is not recorded. These results are not indicative of how well DNA from other species may be sequenced across museums.

Although the relationships among individuals in the expanded intraspecific phylogeny (Fig. 2) is in close agreement with the results reported by Liedigk et al. (2015), we did not find neatly monophyletic lineages based on geography in Clade B, and the divergence dates are slightly earlier than previously inferred (See Table 2). Because the only major difference between our analyses and those of Liedigk et al. is sample size and composition, we believe this may account for the disparity between the two sets of results. Pozzi et al. (2014) demonstrated that extensive taxon sampling can recover phylogenetic relationships that are consistent with morphological and nuclear data in primates. This is consistent with previous studies that raise the concern that incomplete or biased taxon sampling results in phylogenetic error (Hillis et al., 2003; Nabhan and Sarkar, 2012; Plazzi et al., 2010; Townsend and Leuenberger, 2011; Zwickl and Hillis, 2002). So the non-monophyletic relationships we recovered, particularly in Clade B, may be the result of increased and more complete taxon sampling of *M. fascicularis* throughout Southeast Asia. Moreover, various studies indicate that under-sampling taxa in phylogenetic analyses will lead to underestimation of divergence dates, while more complete taxon sampling will recover earlier node ages (Crête-Lafrenière et al., 2012; Milne, 2009; Schulte, 2013), especially when using a relaxed molecular clock model (Soares and Schrago, 2015). This may be the case for this study, as we have more than tripled the sample size of the most recent intraspecific study of *M. fascicularis* (Liedigk et al., 2015).

The inferred intraspecific phylogeny revealed a number of biogeographic patterns as a result of increased taxon sampling. In particular, it emerged from our overall phylogeny based on mitogenomes that the polyphyletic clustering of individuals on Sumatra can now be explained on a geographical basis: Northern Sumatran individuals fall within the mainland cluster (Clade A), whereas southern Sumatran individuals are nested within the insular cluster with populations from Borneo (Clade B). This corroborates the suggestion made by Liedigk et al. (2015) that this could be why Tosi and Coke's (2007) samples from southern Sumatra clustered in a different clade from their own samples from northern Sumatra. Inclusion of museum samples from both the northern and southern ends of Sumatra in this study permitted clear resolution of this issue. Accordingly, Sumatran *M. fascicularis* have two distinct geographic origins. The northern stock most likely originated from the Indo-Chinese Peninsula in the east while the southern lineages appear to be of insular origin, by way of Java and Borneo (through Bangka Island). This is in accord with the Y-chromosomal data from Tosi and Coke (2007), which revealed that both haplogroups are present on Sumatra.

Although this split in populations on Sumatra is rare in vertebrates (Leonard et al., 2015), orangutans also exhibit this phylogeographic pattern intraspecifically on Sumatra (Nater et al., 2011, 2015). Sumatran orangutans (*Pongo abelii*) currently reside only in northern Sumatra, although the subfossil record indicates that their distribution had been widespread across Asia, including the entirety of Sumatra (Delgado and Van Schaik, 2000). This species is particularly diverse in comparison to the other species (*Pongo pygmaeus*) on the island of Borneo (Steiper, 2006). In a series of genetic studies, Nater et al. (2011, 2015) found that one extant population of Sumatran orangutans diverged more recently from *P. pygmaeus* at ~2.09 Ma than from the other populations on Sumatra ~3.50 Ma. This particular population from Batang Toru is the only extant group located to the south of Lake Toba, a large lake that is the site of the recent Toba supereruption approximately 73 kya, along with four other major eruptions within the last 1.2 million years (Chesner et al., 1991). Williams et al. (2009) demonstrated that the supereruption resulted in climatic cooling and prolonged deforestation, which had significant consequences for plants and animals in the region, one of which was to seal off the region north of Lake Toba from the rest of Sumatra (Ambrose, 2003). This boundary at Lake Toba has led to a split that is seen in other taxa. For example, the mountain agile gibbon (*Hylobates agilis*) occurs only south of Lake Toba while the white-handed gibbon (*Hylobates lar*) occurs only to the north (Whittaker et al., 2007; Thinh et al., 2010). The split we see within *M. fascicularis* may also be the result of the Toba eruptions, as the northern Sumatran lineages are north of Lake Toba while the southern Sumatran lineages are well to the south of the lake. While according to our study northern and southern Sumatran lineages diverged approximated 1.88 Ma, which is earlier than the set of major Toba eruptions, it could be possible that the survival of *M. fascicularis* was affected by the Toba supereruption on Sumatra. Once the environment began thriving again, new *M. fascicularis* lineages could migrate to the northern and southern ends of Sumatra via Malaysia and Borneo, respectively, but were still separated by the supereruption site. Although the populations may have met once the forest around Lake Toba was restored after the eruptions, it is unlikely that there would be merging in the mitochondrial genome as *M. fascicularis* are female philopatric (Gerber et al., 2016; Melnick and Hoelzer, 1991; Ruiter and Geffen, 1998). Future studies should examine the nuclear DNA, with a focus on the Y chromosome, to look for patterns of male dispersal in this species.

The hypothesis that the Toba eruption is the cause of the split in the northern and southern Sumatran specimens is, however, questionable because there is no evidence of this division in the mitochondrial DNA of most other Sumatran vertebrates, including rodents, carnivores and birds (Leonard et al., 2015). Leonard et al. (2015) analyzed 28 vertebrate taxa that ranged in body size from rodents to elephants distributed throughout the continental shelf region in Southeast Asia. Sixteen of those taxa had Bornean populations that were basal to Sumatran and Malay Peninsular populations. The divergence dates varied between taxa. All other taxa in their analyses, except for *M. fascicularis*, were either unresolved or had no geographical structure, although this may be due to the fact that only a few mitochondrial genes were analyzed. If the Toba eruption had such a great impact in the distribution of organisms on the Sunda Shelf, it would be expected that many more species, especially terrestrial vertebrates, would display the divide in Sumatra as is seen in long-tailed macaques, orangutans and gibbons.

The majority of the island-living individuals sampled are nested within Clade B, the insular cluster. However, there are four exceptions. The population from Koh Kut, a continental shelf island off the coast of Thailand, is sister to all Thai samples north of the island. The Thai samples that were collected close to the Malaysia-Thailand political border form a sub-cluster within the clade of populations from the Malay Peninsula and northern Sumatra. Another exception is a single individual from the Philippine island of Mindanao (FMNH 56161) that clusters within the populations from northern Thailand. There could be two explanations for this. First, that particular lineage was recently introduced to Mindanao by way of human migration ~2400 years ago. Second, this specimen may have been labeled incorrectly at some point in the past because it is the only specimen out of 12 specimens we had sequenced from Mindanao that nests within Clade A. The final exceptions are the lineages from Bintan and Siantan Islands, both of which are small continental shelf islands off the coast of the Malay Peninsula. These populations likely originated from the mainland stock.

But the majority of insular lineages are clustered within Clade B. As expected, the lineages from islands near Java, such as Bali, Penida and Bawean islands share a common ancestor with lineages from east Java. The Javan specimens form two major groups in Clade B, one of which includes the specimens from Nias and Simeulue Islands to the west of northern Sumatra. This is unexpected as Nias and Simeulue are not geographically close to Java. Additionally, they are considered to be oceanic islands, but whether or not they were connected to the continental shelf during glacial periods in the Pleistocene is unclear. The Nias basin is ~500 m. deep (Deighton et al., 2014), and there do not appear to be any non-volant species that are strictly endemic to the island (Barbour, 1912), so natural colonization of Nias Island may have been possible. The Simeulue basin is deeper at ~1000 m (Milsom, 2005) with very few endemic species, including three snake species, one bird species, one pig species, and a subspecies of *M. fascicularis, M. f. lasiae* (Whitten et al., 2000). Because the strait between these islands and Sumatra is narrow, certain populations may have been able to migrate naturally from the continental shelf by swimming (Gumert and Malaivijitnond, 2012).

The Philippine samples in Clade B appear to have originated from Borneo, which supports previous hypotheses for Philippine colonization by way of Borneo (Abegg and Thierry, 2002; Smith et al., 2014). However, our dating analyses indicate that the Philippines were colonized by long-tailed macaques between ~0.06 and 0.43 Ma, which is much earlier than dating suggested for human-mediated dispersals (Fooden, 2006; Heaney et al., 2016). This indicates that there were multiple waves of migrations likely without human aid. If populations arrived via a land bridge between Borneo and Palawan (Heaney, 1985), then it would be possible for other taxa to cross over to this group of oceanic islands as well, although major faunal differences on either side of the Huxley-Wallace line indicate that this is likely not the case. On the other hand, long-tailed macaques could have swam over and dispersed throughout the oceanic islands via subsequent land bridges amongst these islands. This is a possibility as long-tailed macaques are one of the only non-volant species that is found on either side of this biogeographical divide. Most node support between individuals from the Philippine islands is lacking in statistical significance (Fig. 2). A very recent date of introduction, as is the case with these lineages from the Philippines, could explain the absence of statistically significant node support within certain Philippines macaque groups (Guschanski et al., 2013; Liedigk et al., 2012, 2014; Roos et al., 2011; Zinner et al., 2013b).

In their study using both short tandem repeat (STR) and mtDNA sequences, Smith et al. (2014) suggested that there were at least two paths of dispersal by *M. fascicularis* into the Philippines from Borneo: (1) through Palawan to the northern Philippine islands, and (2) through the Sulu archipelago to the southern Philippine islands. The newly inferred phylogeny indicates that dispersal event (1) through Palawan to the Northern Philippine islands likely led to the distribution of *M. fascicularis* to Luzon in the north but also resulted in the distribution of *M. fascicularis* to some of the more southern islands such as Negros, Bohol, Leyte, and perhaps even Mindanao. The potentially human-mediated dispersal to these southern islands may have been directly through Palawan or through Luzon and Samar (and the adjacent islands). Future studies including samples from Mindoro and Panay Islands would help clarify the dispersal passage to these more southern islands.

Dispersal event (2) of Smith et al. (2014) through the Sulu archipelago to the southern Philippine islands is also consistent with the

results in this study, as individuals from Mindanao cluster with some of the Borneo individuals; but they are monophyletic within a clade that includes an individual from Balabac, an island associated with Palawan. Additionally, the single individual that Liedigk et al. (2015) sequenced from the Philippines appears to be from the southern Philippines, based on its genetic makeup, and clusters with the Mindanao and Borneo individuals from this particular dispersal event. Since we were only able to sequence the mitogenome of one individual from Palawan and no individuals from the islands in the Sulu archipelago—although they do occur in that region (Musser and Heaney, 1985)—future analysis of more individuals from those islands would be useful in determining whether the majority of current lineages in Mindanao had arrived by way of Palawan, through the Sulu archipelago, or both.

In their study, Liedigk et al. (2015) pointed out that the divergence date of approximately 0.93 Ma for the Timor clade that they had sequenced is inconsistent with the hypothesis that lineages to the east of the Wallace line were introduced within the last 4000 years (Fooden, 2006; Heaney et al., 2016). However, based on genetic data, modern humans had colonized the Timor region by 37 ka (Gomes et al., 2015), so *M. fascicularis* could have been introduced by humans much earlier. The present study estimates an even earlier divergence date of Timor individuals from all other insular individuals of approximately 1.03 Ma for this clade from the rest of Clade B, although the most recent common ancestor within the Timor clade is estimated to be 0.03 Ma. The divergence date and placement of this clade in the phylogeny implies an unexpected history for this lineage. Liedigk et al. (2015) suggested that the Timor lineage was the sister group in Clade B because the Timor haplotypes had originated from an area somewhere in Sundaland that they had not sampled in their study. However, our thorough sampling throughout Sundaland (Fig. 2) does not suggest a locality for the origin of the Timor clade. Still, this does not eliminate the possibility that the Timor lineage had naturally colonized the island. Future studies that include specimens from islands near Timor, such as Lombok, Sumbawa, Flores, and Sumba may help illuminate the biogeographic history of this clade.

It must be borne in mind that the inferred phylogeny is based on mtDNA alone. It therefore reflects population-level cladogenic events attributable to the sedentary nature of female macaques (de Ruiter and Geffen, 1998; Pusey and Packer, 1987) and cannot reveal any influences exerted by the typical wide dispersal of males. Additionally, the mitogenome is a single marker, so incomplete lineage sorting can lead to a discrepancy between the phylogeny reconstructed in the study and the species tree (Roger and Gibbs, 2014). Analyses of the nuclear genome could help further clarify the dispersal of *M. fascicularis* throughout Southeast Asia and would reveal at least some of the genetic selection throughout the colonization of this region.

5. Conclusion

In this study, we have elucidated biogeographic patterns of *M. fascicularis* throughout Southeast Asia by using thorough sampling of museum specimens from the region, and in the process, we have identified avenues that still need to be pursued in order to fully understand this widespread species: It is necessary to examine the nuclear DNA, especially the Y chromosome, to determine patterns of male dispersal in this species. We were not able to determine the origins of the Timor specimens, and including specimens from islands near Timor, such as Lombok, Sumbawa, Flores, and Sumba could illuminate the biogeographic history of this clade. Although we have sampled widely throughout the Philippines, it appears that there may be two major migration routes into this region. Analyzing samples from Mindoro, Panay, Palawan, and the islands in the Sulu archipelago would help determine dispersal passages through this chain of oceanic islands.

We have shown here that the use of next-generation sequencing, particularly shotgun sequencing, on museum specimens proves to be a successful way of sequencing entire mitogenomes at high coverage due to the nature of ancient DNA. The newly reconstructed intraspecific phylogeny demonstrates that lineages within *M. fascicularis* diverged earlier than previously inferred, that there is a separation in Sumatran macaques, and that there is a colonization pattern in the Philippines. The results of this study demonstrate the importance of extensive geographic sampling for a widespread species for understanding biogeographic diversification in Southeast Asia because it expands knowledge about the relationships and divergence dates throughout the entire region rather than the limited geographic ranges for most taxa. *M. fascicularis* is commonly studied for both evolutionary and biomedical purposes, and we hope these findings will be of use to future studies that use this species as a model as well.

Acknowledgements

We gratefully acknowledge kind cooperation in providing access to their respective museum collections for tissue sampling from the late William Stanley and Dr. Lawrence Heaney at The Field Museum of Natural History (FMNH), Eileen Westwig and Neil Duncan at the American Museum of Natural History (AMNH), Darrin P. Lunde at the Smithsonian Institution National Museum of Natural History (NMNH) and Pepijn Kamminga and Steven van der Mije at the Naturalis Biodiversity Center (Leiden, Netherlands). We are also grateful to Dr. Lawrence Heaney for his advice on Southeast Asian biogeography, to Dr. Callum F. Ross for providing us with fresh tissues of *M. fascicularis*, to John Lindo, Elizabeth Mallott, Kelsey Witt and Amanda Owings for their consultation on the ancient DNA mitogenome pipeline, and to Alvaro Gonzalo Hernandez and Chris Wright for their guidance with next generation sequencing. Finally, we would like to thank Dr. Lawrence Heaney, Dr. Callum F. Ross, Dr. ZheXi Luo, Dr. Christian Roos, Dr. Luca Pozzi, Dr. Todd Disotell and Dr. James Herrera for providing comments or advice on this manuscript.

Funding

This research was supported by the NSF DDIG DEB-1501733 to Robert D. Martin and Lu Yao, grants to Lu Yao (NSF GRFP 100152161, The Field Museum Women-in-Science Graduate Fellowship, Arts and Science Fund from the Arts, Science and Culture Initiative at University of Chicago, and Hinds Fund Award, Field Museum Pritzker Lab Award, and Travel Award from the Committee on Evolutionary Biology at University of Chicago), research funding from The Field Museum to Robert D. Martin and research funds from the University of Illinois Urbana-Champaign to Ripan S. Malhi.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ympev.2017.08.006.

References

Abegg, C., Thierry, B., 2002. Macaque evolution and dispersal in insular south-east Asia. Biol. J. Linn. Soc. 75, 555–576.

Aimi, M., Aziz, F., 1985. Vertebrate fossils from the Sangiran Dome, Mojokerto, Trinil and Sambungmacam areas. In: Watanabe, N., Kadar, D. (Eds.), Quaternary Geology of the Hominid Fossil Bearing Formations in Java, Report of the Indonesia-Japan Joint Research Project CTA-41, 1976–1979. Geological Research and Development Centre, Bandung, pp. 155–198.

Alba, D.M., Delson, E., Carnevale, G., Colobero, S., Delfino, M., Giuntelli, P., Pavia, M., Pavia, G., 2014. First joint record of *Mesopithecus* and cf. *Macaca* in the Miocene of Europe. J. Hum. Evol. 67, 1–18.

Ambrose, S.H., 2003. Did the super-eruption of Toba cause a human population bottleneck? Reply to Gathorne-Hardy and Harcourt-Smith. J. Hum. Evol. 45, 231–237.

Bailey, S.E., Mao, X., Bozek, M., Tsagkogeorga, G., Csorba, G., Heaney, L.R., Sedlock, J., Stanley, W., Rouillard, J.-M., Rossiter, S.J., 2015. Successful targeted sequence capture of degraded museum samples—a phylogenomic study of horseshoe bats. Biol. J. Linnean Soc. 117, 58–70.

Barbour, T., 1912. A Contribution to the Zoögeography of the East Indian Islands. Museum of Comparative Zoölogy at Harvard College, Cambridge.

Bergman, S.C., Coffield, D.Q., Talbot, J.P., Garrard, R.A., 1996. Tertiary Tectonic and magmatic evolution of western Sulawesi and the Makassar Strait, Indonesia: evidence for a Miocene continent-continent collision. In: Hall, R., Blundell, D. (Eds.), Tectonic Evolution of Southeast Asia. Geological Society Special Publication No. 106, pp. 391–429.
Bonhomme, M., Cuartero, S., Blancher, A., Crouau-Roy, B., 2009. Assessing natural introgression in 2 biomedical model species, the rhesus macaque (*Macaca mulatta*) and the long-tailed macaque (*Macaca fascicularis*). J. Hered. 100, 158–159.
Briggs, A.W., Good, J.M., Green, R.E., Krause, J., Maricic, T., Stenzel, U., Lalueza-Fox, C., Rudan, P., Brajkovic, D., Kucan, Z., Gusic, Íl., Schmitz, R., Doronichev, V.B., Golovanova, L.V., De la Rasilla, M., Fortea, J., Rosas, A., Pääbo, S., 2009. Targeted retrieval and analysis of five Neandertal mtDNA genomes. Science 325, 318–321.
Brown, W.M., George Jr., M., Wilson, A.C., 1979. Rapid evolution of animal mitochondrial DNA. Proc. Natl. Acad. Sci. USA 76, 1967–1971.
Brunet, M., Guy, F., Pilbeam, D., Lieberman, D.E., Likius, A., Mackaye, H.T., Ponce de León, M.S., Zollikofer, C.P.E., Vignaud, P., 2005. New material of the earliest hominid from the Upper Miocene of Chad. Nature 434, 752–755.
Burrell, A.S., Disotell, T.R., Bergey, C.M., 2015. The use of museum specimens with high-throughput DNA sequencers. J. Hum. Evol. 79, 35–44.
Chan, Y., Roos, C., Inoue-Murayama, M., Inoue, E., Shih, C., Pei, K.J., Vigilant, L., 2010. Mitochondrial genome sequences effectively reveal the phylogeny of *Hylobates* gibbons. PLoS ONE 5 (12), e144919.
Chatterjee, H.J., 2006. Phylogeny and biogeography of gibbons: a dispersal-vicariance analysis. Int. J. Primatol. 27, 699–712.
Chesner, C.A., Rose, W.I., Deino, A., Drake, R., Westgate, J.A., 1991. Eruptive history of Earths largest Quaternary caldera (Toba, Indonesia) clarified. Geology 19, 200–203.
Crête-Lafrenière, A., Weir, L.K., Bernatchez, L., 2012. Framing the Salmonidae Family phylogenetic portrait: a more complete picture from increased taxon sampling. PLoS ONE 7, e466662.
Cui, Y., Lindo, J., Hughes, C.E., Johnson, J.W., Hernandez, A.G., Kemp, B.M., Ma, J., Cunningham, R., Petzelt, B., Mitchell, J., Archer, D., Cybulski, J.S., Malhi, R.S., 2013. Ancient DNA analysis of Mid-Holocene individuals from the Northwest coast of North America reveals different evolutionary paths for mitogenomes. PLoS ONE 8, e66948.
Deighton, I., Mukti, M.M., Singh, S., Travis, T., Harwick, A., Hernon, K., 2014. Nias Basin, NW Sumatra – new insights into forearc structure and hydrocarbon prospectivity from long-offset 2D seismic data. Proc., Indonesian Petroleum Association, 38th Annual Convention and Exhibition.
Delgado, R.A., Van Schaik, C.P., 2000. The behavioral ecology and conservation of the orangutan (*Pongo pygmaeus*): a tale of two islands. Evol. Anthropol. 9, 201–218.
Delson, E., 1980. Fossil macaques, phyletic relationships and a scenario of deployment. In: Lindburg, D.G. (Ed.), The Macaques: Studies in Ecology, Behavior, and Evolution. Van Nostrand Reinhold, New York, pp. 10–30.
Delson, E., 2000. Cercopithecinae. In: Delson, E., Tattersall, I., Van Couvering, J.A., Brooks, A.S. (Eds.), Encyclopedia of Human Evolution and Prehistory. Garland Publishing Inc., New York, pp. 166–171.
de Ruiter, J.R., Geffen, E., 1998. Relatedness of matrilines, dispersing males and social groups in long-tailed macaques (*Macaca fascicularis*). Proc. Biol. Sci. 265, 79–87.
Drummond, A.J., Ho, S.Y., Phillips, M.J., Rambaut, A., 2006. Relaxed phylogenetics and dating with confidence. PLoS Biol. 4, e88.
Drummond, A.J., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. 7, 214.
Evans, A.R., Jones, D., Boyer, A.G., Brown, J.H., Costa, D.P., Ernest, S.K.M., Fitzgerald, E.M.G., Fortelius, M., Gittleman, J.L., Hamilton, M.J., Harding, L.E., Lintulaakso, K., Lyons, S.K., Okie, J.G., Saarinen, J.J., Sibly, R.M., Smith, F.A., Stephens, P.R., Theodor, J.M., Uhen, M.D., 2012. The maximum rate of mammal evolution. Proc. Natl. Acad. Sci. USA 109, 4027–4028.
Fooden, J., 1964. Rhesus and crab-eating macaques: intergradation in Thailand. Science 143, 363–364.
Fooden, J., 1976. Provisional classification and key to living species of macaques (Primates: *Macaca*). Folia Primatol. 25, 225–236.
Fooden, J., 1995. Systematic review of Southeast Asian longtail macaques, *Macaca fascicularis* (Raffles, [1821]). Fieldiana Zool. 81, 1–228.
Fooden, J., 2000. Systematic review of the rhesus macaque, *Macaca mulatta* (Zimmermann, 1780). Fieldiana Zool. 96, 1–196.
Fooden, J., 2006. Comparative review of fasicularis-group species of macaques (Primates: *Macaca*). Fieldiana Zool. 107, 1–43.
George, W., 1981. Wallace and his line. In: Whitmore, T.C. (Ed.), Wallace's Line and Plate Tectonics. Clarendon Press, Oxford, pp. 3–8.
Gerber, L., Krützen, M., de Ruiter, J.R., van Schaik, C.P., van Noordwijk, M.A., 2016. Postdispersal nepotism in male long-tailed macaques (*Macaca fascicularis*). Ecol. Evol. 6, 46–55.
Ginolhac, A., Rasmussen, M., Thomas, M., Willerslev, E., Orlando, L., 2011. MapDamage: testing for damage patterns in ancient DNA sequences. Bioinformatics 27, 2153–2155.
Gomes, S.M., Bodner, M., Souto, L., Zimmermann, B., Huber, G., Strobl, C., Röck, A.W., Achilli, A., Olivieri, A., Torroni, A., Côte-Real, F., Parson, W., 2015. Human settlement history between Sunda and Sahul: a focus on East Timor (Timor-Leste) and the Pleistocenic mtDNA diversity. BMC Genom. 16, 70.
Groves, C.P., 2001. Primate Taxonomy. Smithsonian Institution Press, Washington.
Gumert, M.D., Malaivijitnond, S., 2012. Marine prey processed with stone tools by Burmese long-tailed macaques (*Macaca fascicularis aurea*) in intertidal habitats. Am. J. Phys., Anthropol. 149, 447–457.
Guschanski, K., Krause, J., Sawyer, S., Valente, L.M., Bailey, S., Finstermeier, K., Sabin, R., Gilissen, E., Sonet, G., Nagy, Z.T., Lenglet, G., Mayer, F., Savolainen, V., 2013. Next-generation museomics disentangle one of the largest primate radiations. Syst. Biol. 62, 539–554.
Hagelberg, E., Clegg, J.B., 1991. Isolation and characterization of DNA from archaeological bone. Proc. Biol. Sci. 244, 45–50.
Heaney, L.R., Balete, D.S., Rickart, E.A., 2016. The Mammals of Luzon Island: Biogeography and Natural History of a Philippine Fauna. Johns Hopkins University Press, Baltimore.
Heaney, L.R., 1986. Biogeography of mammals in SE Asia: estimates of rates of colonization, extinction and speciation. Biol. J. Linn. Soc. 28, 127–165.
Heaney, L.R., 1985. Zoogeographic evidence for Middle and Late Pleistocene land bridges to the Philippine islands. Mod. Quatern. Res. S.E. Asia 9, 127–143.
Higuchi, R., Bowman, B., Freiberger, M., Ryder, O.A., Wilson, A.C., 1984. DNA sequences from the quagga, an extinct member of the horse family. Nature 312, 282–284.
Hillis, D.M., Pollock, D.D., McGuire, J.A., Zwickl, D.J., 2003. Is sparse taxon sampling a problem for phylogenetic inference? Syst. Biol. 52, 124–126.
Huxley, T.H., 1868. On the classification and distribution of the Alectoromorphae and Heteromorphae. Proc. Zool. Soc. Lond. 1868, 294–319.
Jansa, S.A., Barker, F.K., Heaney, L.R., 2006. The pattern and timing of diversification of Philippine endemic rodents: evidence from mitochondrial and nuclear gene sequences. Syst. Biol. 55, 73–88.
Kanthaswamy, S., Satkoski, J., George, D., Kou, A., Erickson, B.J.-A., Smith, D.G., 2009. Interspecies hybridization and the stratification of nuclear genetic variation of rhesus (*Macaca mulatta*) and long-tailed macaques (*Macaca fascicularis*). Int. J. Primatol. 29, 1295–1311.
Katoh, S., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30, 772–780.
Kelley, J., 2002. The hominoid radiation in Asia. In: Hartwig, W.C. (Ed.), The Primate Fossil Record. Cambridge University Press, Cambridge, pp. 369–384.
Kircher, M., Sawyer, S., Meyer, M., 2012. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. Nucl. Acids Res. 40, e3.
Krause, J., Fu, Q., Good, J.M., Viola, B., Shunkov, M.V., Derevianko, A.P., Pääbo, S., 2010. The complete mitochondrial DNA genome of an unknown hominin from southern Siberia. Nature 464, 894–897.
Lanfear, R., Calcott, B., Ho, S.Y.W., Guindon, S., 2012. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Mol. Biol. Evol. 29, 1695–1701.
Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359.
Leakey, M.G., 1993. Evolution of *Theropithecus* in the Turkana Basin. In: Jablonski, N.G. (Ed.), *Theropithecus*, the Rise and Fall of a Primate Genus. Cambridge University Press, Cambridge, pp. 85–124.
Lebatard, A.E., Bourlès, D.L., Duringer, P., Jolivet, M., Braucher, R., Carcaillet, J., Schuster, M., Arnaud, N., Monié, P., Lihoreau, F., Likius, A., Mackaye, H.T., Vignaud, P., Brunet, M., 2008. Cosmogenic nuclide dating of *Sahelanthropus tchadensis* and *Australopithecus bahrelghazali*: Mio-Pliocene hominids from Chad. Proc. Natl. Acad. Sci. USA 105, 3226–3231.
Leonard, J.A., den Tex, R.-J., Hawkins, M.T.R., Muñoz-Fuentes, V., Thorington, R., Maldonado, J.E., 2015. Phylogeography of vertebrates on the Sunda Shelf: a multi-species comparison. J. Biogeogr. 42, 871–879.
Li, J., Han, K., Xing, J., Kim, H.S., Rogers, J., Ryder, O.A., Disotell, T., Yue, B., Batzer, M.A., 2009. Phylogeny of the macaques (Cercopithecidae: *Macaca*) based on Alu elements. Gene 488, 242–249.
Liedigk, R., Kolleck, J., Böker, K.O., Meijaard, E., Md-Zain, B.M., Abdul-Latiff, M.A.B., Ampeng, A., Lakim, M., Abdul-Patah, P., Tosi, A.J., Brameier, M., Zinner, D., Roos, C., 2015. Mitogenomic phylogeny of the common long-tailed macaque (*Macaca fascicularis fascicularis*). BMC Genom. 16, 222.
Liedigk, R., Roos, C., Brameier, M., Zinner, D., 2014. Mitogenomics of the Old World monkey tribe Papionini. BMC Evol. Biol. 14, 176.
Liedigk, R., Yang, M., Jablonski, N.G., Momberg, F., Geissmann, T., Lwin, N., Hla, T.H., Liu, Z., Wong, B., Ming, L., Yongcheng, L., Zhang, Y.-P., Nadler, T., Zinner, D., Roos, C., 2012. Evolutionary history of the odd-nosed monkeys and the phylogenetic position of the newly described Myanmar snub-nosed monkey *Rhinopithecus strykeri*. PLoS ONE 7, e37418.
Lindo, J., Huerta-Sanchez, E., Nakgome, S., Rasmussen, M., Petzelt, B., Mitchell, J., Cybulski, J.S., Willerslev, E., DeGiorgio, M., Malhi, R.S., 2016. A time transect of exomes from a Native American population before and after European contact. Nat. Commun. 7, 13175.
Maddison, W.P., Maddison, D.R., 2011. Mesquite: A Modular System for Evolutionary Analysis. Version 2.75. < http://mesquiteproject.org > .
Malström, H., Svensson, E., Gilbert, T., Willerslev, E., Götherström, A., Holmlund, G., 2007. More on contamination: the use of the asymmetric molecular behavior to identify authentic ancient human DNA. Mol. Biol. Evol. 24, 998–1004.
Mason, V.C., Li, G., Helgen, K.M., Murphy, W.J., 2011. Efficient cross-species capture hybridization and next-generation sequencing of mitochondrial genomes from non-invasively sampled museum specimens. Genome Res. 21, 1695–1704.
Melnick, D.J., Hoelzer, G.A., 1991. Differences in male and female macaque dispersal lead to contrasting distribution of nuclear and mitochondrial DNA variation. Int. J. Primatol. 13, 379–393.
Miller, M.A., Pfeiffer, W., Schwartz, T., 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: Proceedings of the Gateway Computing Environments Workshop (GCE). New Orleans, pp. 1–8.
Milne, R.I., 2009. Effects of taxon sampling on molecular dating for within-genus divergence events, when deep fossils are used for calibration. J. Syst. Evol. 47, 383–401.
Milsom, J., 2005. Seismology and neotectonics. In: Barber, A.J., Crow, M.J., Milsom, J.S. (Eds.), Sumatra: Geology, Resources and Tectonic Evolution. The Alden Press, Oxford, pp. 8–15.
Musser, G.G., Heaney, L.R., 1985. Philippine *Rattus*: a new species from the Sulu Archipelago. Am. Mus. Novit. 2818, 1–32.

Nabhan, A.R., Sarkar, I.N., 2012. The impact of taxon sampling on phylogenetic inference: a review of two decades of controversy. Brief. Bioinform. 13, 122–134.
Nater, A., Greminger, M.P., Arora, N., vanSchaik, C.P., Goossens, B., Singleton, I., Verschoor, E.J., Warren, K.S., Krützen, M., 2015. Reconstructing the demographic history of orang-utans using Approximate Bayesian Computation. Mol. Ecol. 24, 310–327.
Nater, A., Nietlisbach, P., Arora, N., van Schaik, C.P., van Noordwijk, M.A., Willems, E.P., Singleton, I., Wich, S.A., Goossens, B., Warren, K.S., Verschoor, E.J., Perwitasari-Farajallah, D., Pamungkas, J., Krützen, M., 2011. Sex-biased dispersal and volcanic activities shaped phylogeographic patterns of extant orangutans (genus: *Pongo*). Mol. Biol. Evol. 28, 2275–2288.
Ong, P., Richardson, M., 2008. Macaca fascicularis. The IUCN Red List of Threatened Species. eT12551A3355536.
Outlaw, D.C., Voelker, G., 2008. Pliocene climatic change in insular Southeast Asia as an engine of diversification in *Ficedula* flycatchers. J. Biogeogr. 35, 739–752.
Pääbo, S., 1989. Ancient DNA: extraction, characterization, molecular cloning, and enzymatic amplification. Proc. Natl. Acad. Sci. USA 86, 1939–1943.
Pääbo, S., Higuchi, R.G., Wilson, A.C., 1989. Ancient DNA and the polymerase chain reaction. J. Biol. Chem. 264, 9709–9712.
Pacheco, M.A., Battistuzzi, F.U., Lentino, M., Aguilar, R.F., Kumar, S., Escalante, A.A., 2011. Evolution of modern birds revealed by mitogenomics: timing the radiation and origin of major orders. Mol. Biol. Evol. 28, 1927–1942.
Plazzi, F., Ferrucci, R.R., Passamonti, M., 2010. Phylogenetic representativeness: a new method for evaluating taxon sampling in evolutionary studies. BMC Bioinform. 11, 209.
Pozzi, L., Hodgson, J.A., Burrell, A.S., Disotell, T.R., 2011. The stem catarrhine *Saadanius* does not inform the timing of the origin of crown catarrhines. J. Hum. Evol. 61, 209–210.
Pozzi, L., Hodgson, J.A., Burrell, A.S., Sterner, K.N., Raaum, R.L., Disotell, T.R., 2014. Primate phylogenetic relationships and divergence dates inferred from complete mitochondrial genomes. Mol. Phylogenet. Evol. 75, 165–183.
Pusey, A.E., Packer, C., 1987. Dispersal and philopatry. In: Smuts, B.B., Cheney, D.L., Seyfarth, R.M., Wrangham, R.W., Struhsaker, T.T. (Eds.), Primate Societies. University of Chicago Press, Chicago, pp. 250–266.
Rambaut, A., FigTree: Tree Figure Drawing Tool, Version 1.4.2. < http://tree.bio.ed.ac.uk/software/figtree/ > (accessed 10.01.16).
Roger, J., Gibbs, R.A., 2014. Comparative primate genomics: emerging patterns of genome content and dynamics. Nat. Rev. Genet. 15, 347–359.
Rohland, N., Malaspinas, A.S., Pollack, J.L., Slatkin, M., Matheus, P., Hofreiter, M., 2007. Proboscidean mitogenomics: chronology and mode of elephant evolution using mastodon as outgroup. PLoS Biol. 5, e207.
Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574.
Roos, C., Zinner, D., Kubatko, L.S., Schwarz, C., Yang, M., Meyer, D., Nash, S.D., Xing, J., Batzer, M.A., Brameier, M., Leendertz, F.H., Ziegler, T., Perwitasari-Farajallah, D., Nadler, T., Walter, L., Osterholz, M., 2011. Nuclear versus mitochondrial DNA: evidence for hybridization in colobine monkeys. BMC Evol. Biol. 11, 77.
Rowe, K.C., Singhal, S., Macmanes, M.D., Ayroles, J.F., Morelli, T.L., Rubidge, E.M., Bi, K., Moritz, C.C., 2011. Museum genomics: low-cost and high-accuracy genetic data from historical specimens. Mol. Ecol. Resourc. 11, 1082–1092.
Ruiter, J.R.D., Geffen, E., 1998. Relatedness of matrilines, dispersing males and social groups in long-tailed macaques (*Macaca fascicularis*). Proc. R. Soc. Lond. B 265, 79–87.
Schulte, J.A., 2013. Undersampling taxa will underestimate molecular divergence dates: an example from the South American Lizard Clade Liolaemini. Int. J. Evol. Biol. 2013, 628467.
Smith, D.G., McDonough, J.W., George, D.A., 2007. Mitochondrial DNA variation within and among regional populations of long-tail macaques (*Macaca fascicularis*) in relation to other species of the *fascicularis* group of macaques. Am. J. Primatol. 69, 182–198.
Smith, D.G., Ng, J., George, D., Trask, J.S., Houghton, P., Singh, B., Villano, J., Kanthaswamy, S., 2014. A genetic comparison of two alleged subspecies of Philippine cynomolgus macaques. Am. J. Phys. Anthropol. 155, 136–148.
Soares, A.E., Schrago, C.G., 2015. The influence of taxon sampling on Bayesian divergence time inference under scenarios of rate heterogeneity among lineages. J. Theor. Biol. 364, 31–39.
Stamatakis, A., 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22, 2688–2690.
Steiper, M.E., 2006. Population history, biogeography, and taxonomy of orangutans (Genus: *Pongo*) based on a population genetic meta-analysis of multiple loci. J. Hum. Evol. 50, 509–522.
Stevens, N.J., Seiffert, E.R., O'Connor, P.M., Roberts, E.M., Schmitz, M.D., Krause, C., Gorscak, E., Ngasala, S., Hieronymus, T.L., Temu, J., 2013. Palaeontological evidence for an Oligocene divergence between Old World monkeys and apes. Nature 497, 611–614.
Stevison, L.S., Kohn, M.H., 2009. Divergence population genetic analysis of hybridization between rhesus and cynomolgus macaques. Mol. Ecol. 18, 2457–2475.
Steppan, S.J., Zawadzki, C., Heaney, L.R., 2003. Molecular phylogeny of the endemic Philippine rodent *Apomys* (Muridae) and the dynamics of diversification in an oceanic archipelago. Biol. J. Linn. Soc. 80, 699–715.
Suchard, M.A., Rambaut, A., 2009. Many-core algorithms for statistical phylogenetics. Bioinformatics 25, 1370–1376.
Swofford, D.L., 2002. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.
Thinh, V.N., Mootnick, A.R., Geissmann, T., Li, M., Ziegler, T., Agil, M., Moisson, P., Tilo, N., Walter, L., Roos, C., 2010. Mitochondrial evidence for multiple radiations in the evolutionary history of small apes. BMC Evol. Biol. 10, 74.
Tosi, A.J., Coke, C.S., 2007. Comparative phylogenetics offer new insights into the biogeographic history of *Macaca fascicularis* and the origin of the Mauritian macaques. Mol. Phylogenet. Evol. 42, 498–504.
Tosi, A.J., Morales, J.C., Melnick, D.J., 2003. Paternal, maternal, and biparental molecular markers provide unique windows onto the evolutionary history of macaque monkeys. Evolution 57, 1419–1435.
Tosi, A.J., Morales, J.C., Melnick, D.J., 2002. Y-chromosome and mitochondrial markers in *Macaca fascicularis* indicate introgression with Indochinese *M. mulatta* and a biogeographic barrier in the Isthmus of Kra. Int. J. Primatol. 23, 161–178.
Townsend, J.P., Leuenberger, C., 2011. Taxon sampling and the optimal rates of evolution for phylogenetic inference. Syst. Biol. 60, 358–365.
Vignaud, P., Duringer, P., Mackaye, H.T., Likius, A., Blondel, C., Boisserie, J.R., de Bonis, L., Eisenmann, V., Etienne, M.E., Geraads, D., Guy, F., Lehmann, T., Lihoreau, F., Lopez-Martinez, N., Mourer-Chauviré, C., Otero, O., Rage, J.C., Schuster, M., Viriot, L., Zazzo, A., Brunet, M., 2002. Geology and palaeontology of the Upper Miocene Toros-Menalla hominid locality, Chad. Nature 418, 152–155.
Wallace, A.R., 1863. On the physical geography of the Malay Archipelago. J. Roy. Geog. Soc. 7, 205–212.
Whittaker, D.J., Morales, J.C., Melnick, D.J., 2007. Resolution of the *Hylobates* phylogeny: congruence of mitochondrial D-loop sequences with molecular, behavioral, and morphological data sets. Mol. Phylogenet. Evol. 45, 620–628.
Whitten, T., Damanik, S.J., Anwar, J., Hisyam, N., 2000. The Ecology of Sumatra. Tuttle Publishing, North Clarendon.
Williams, M.A.J., Ambrose, S.H., van der Kaars, S., Ruehlemann, C., Chattopadhyaya, U., Pal, J., Chauhan, P.R., 2009. Environmental impact of the 73 ka Toba super-eruption in South Asia. Palaeogeog. Palaeoclimatol. Palaeoecol. 284, 295–314.
Yan, G., Zhang, G., Fang, X., Zhang, Y., Li, C., Ling, F., Cooper, D.N., Li, Q., Li, Y., van Gool, A.J., Du, H., Chen, J., Chen, R., Zhang, P., Huang, Z., Thompson, J.R., Meng, Y., Bai, Y., Wang, J., Zhuo, M., Wang, T., Huang, Y., Wei, L., Li, J., Wang, Z., Hu, H., Yang, P., Le, L., Stenson, P.D., Li, B., Liu, X., Ball, E.V., An, N., Huang, Q., Zhang, Y., Fan, W., Zhang, X., Li, Y., Wang, W., Katze, M.G., Su, B., Nielsen, R., Yang, H., Wang, J., Wang, X., Wang, J., 2011. Genome sequencing and comparison of two nonhuman primate animal models, the cynomolgus and Chinese rhesus macaque. Nat. Biotechnol. 29, 1019–1023.
Zalmout, I.S., Sanders, W.J., MacLatchy, L.M., Gunnell, G.F., Al-Mufarreh, Y.A., Ali, M.A., Nasser, A.A.H., Al-Masari, A.M., Al-Sobhi, S.A., Nadhra, A.O., Matari, A.D., Wilson, J.A., Gingerich, P.D., 2010. New Oligocene primate from Saudi Arabia and the divergence of apes and Old World monkeys. Nature 466, 360–365.
Zinner, D., Fickenscher, G.H., Roos, C., 2013a. Family Cercopithecidae (Old World Monkeys). In: Mittermeier, R.A., Rylands, A.B., Wilson, D.E. (Eds.), Handbook of the Mammals of the World. Primates, vol. 3. Lynx Edicions, Barcelona, pp. 550–627.
Zinner, D., Wertheimer, J., Liedigk, R., Groeneveld, L.F., Roos, C., 2013b. Baboon phylogeny as inferred from complete mitochondrial genomes. Am. J. Phys. Anthropol. 150, 133–140.
Zwickl, D.J., Hillis, D.M., 2002. Increased taxon sampling greatly reduces phylogenetic error. Syst. Biol. 51, 588–598.