

Evolutionary pattern of *Macaca fascicularis* in Southeast Asia inferred using Y-chromosomal gene

Jeffrine J Rovie-Ryan

Ex-Situ Conservation Division <https://orcid.org/0000-0003-2910-5431>

Faisal Ali Anwarali Khan (✉ akfali@unimas.my)

Universiti Malaysia Sarawak (UNIMAS) <https://orcid.org/0000-0001-7585-4885>

Mohd Tajuddin Abdullah

Akademi Sains Malaysia

Research article

Keywords: TSPY, SRY, Indochinese, continental and insular Sundaic, hybridization, secondary contact

Posted Date: January 8th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-49299/v2>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Version of Record: A version of this preprint was published on February 15th, 2021. See the published version at <https://doi.org/10.1186/s12862-021-01757-1>.

Abstract

Background: We analyzed a combined segment (2,032-bp) of the sex-determining region and the testis-specific protein of the Y-chromosome (Y-DNA) gene to clarify the gene flow and phylogenetic relationships of the long-tailed macaques (*Macaca fascicularis*) in Southeast Asia. Phylogenetic relationships were constructed using the maximum likelihood, Bayesian inference, and the median-joining network from a total of 164 adult male *M. fascicularis* from 62 localities in Malaysia, including sequences from the other regions from previous studies.

Results: Based on Y-DNA, we confirm the presence of two lineages of *M. fascicularis*: the Indochinese and Sundaic lineages. The Indochinese lineage is represented by *M. fascicularis* located northwards of the Surat Thani-Krabi depression region and is introgressed by the *Macaca mulatta* Y-DNA. The Sundaic lineage is free from such hybridization event, thus defined as the original carrier of the *M. fascicularis* Y-DNA. We further revealed that the Sundaic lineage differentiated into two forms: the insular and the continental forms. The insular form, which represents the ancestral form of *M. fascicularis*, consists of two haplotypes: a single homogenous haplotype occupying the island of Borneo, Philippines, and southern Sumatra; and the Javan haplotype. The more diverse continental form consists of 17 haplotypes in which a dominant haplotype was shared by individuals from southern Thai Peninsular (south of Surat Thani-Krabi depression), Peninsular Malaysia, and Sumatra. Uniquely, Sumatra contains both the continental and insular Y-DNA which can be explained by a secondary contact hypothesis.

Conclusions: Overall, the findings in this study are important: 1) to help authority particularly in Malaysia on the population management activities including translocation and culling of conflict *M. fascicularis*, 2) to identify the unknown origin of captive *M. fascicularis* used in biomedical research, and; 3) the separation between the continental and insular forms warrants for the treatment as separate management units.

Background

The most common and abundant *Macaca* species in Malaysia, *Macaca fascicularis* (Raffles, 1821) (also known as long-tailed macaques, crab-eating macaques, and cynomolgus macaques), are widely distributed in nature and occupy vast areas of mainland Southeast Asia (SEA) (Bangladesh, Thailand, Cambodia, Vietnam, Laos, Myanmar, Peninsular Malaysia and Singapore), the Greater and Lesser Sunda Islands (Indonesia, Brunei, and the Malaysian Borneo), and the Philippines, extending across 30° of latitude and 35° of longitude [1]. This species is highly adaptable to disturbed environments and secondary forests, and commonly populate low elevation habitat types favouring seashores and mangrove forests, riverbanks, and swamp forests [2,3]. In Peninsular Malaysia, they are widespread and populate areas in sympatry with human settlements [4,5]. At the same time, in the Malaysian Borneo (Sabah and Sarawak), they are distributed throughout the lowlands populating the coastal regions [6].

M. fascicularis like all the other species within the genus *Macaca* demonstrates a high level of male-mediated dispersal while females are sedentary in nature [7,8]. This implies that the transfer of genes between groups and populations occurs almost exclusively through the dispersal of males [8]. Moreover, the dominant male exhibits a polygynous mating system, which results in a homogenous gene pool of offspring. Because of this characteristics, particular *M. fascicularis* groups or populations would consist of homogenized Y-chromosomal DNA (Y-DNA) [9].

Y-DNA is uniparentally inherited and as a haploid genetic system, is passed down paternally from male adults to their male progenies. Due to the lack of recombination, Y-DNA has one-fourth of the effective population size (N_e) of that of nuclear DNA [9, 10]. Additionally, the polygynous mating behaviour of dominant males further lowers the N_e of Y-DNA loci [9,11-14]. Nevertheless, the paternal inheritance and the highly dispersal nature of male macaques mean that the information from Y-DNA is important to infer the male-mediated gene flow and to distinguish between contemporary gene flow and ancestral polymorphism [15].

The sex-determining region (SRY) and the testis-specific protein (TSPY) are two unlinked gene segments within the Y-DNA which have been used to infer male-mediated gene flow in primates, particularly in Cercopithecinae [9,15-23]. Specifically, Tosi, Morales and Melnick [15] and Bunlungsup et al. [23] used both the SRY and TSPY segments coupled with mitochondrial (mtDNA) to investigate the contact zone between *M. fascicularis* and *M. mulatta* in the Indochinese region while Tosi, Morales, and Melnick [9,16] used a small number of representatives from various region of SEA to investigate the phylogeography of *M. fascicularis*.

The widespread distribution of *M. fascicularis* makes it a prime candidate for the study of phylogeography and evolutionary studies. During the Quaternary Periods (Pleistocene and Holocene), the currently separated landmasses in SEA were intermittently joined and

formed Sundaland that permits the dispersal of land mammals, including primates [24-26]. The climatic oscillations, especially during the Pleistocene, including major volcanic eruptions of Mount Toba, would have induced vegetation changes causing the expansions or contractions of habitat that influenced the dispersal of animals [27-29]. Moreover, major river systems, lakes, and mountain ranges may also have acted as a barrier to gene flow [26,28].

In Peninsular Malaysia, the rising number of cases of human-*Macaca* conflict, particularly involving *M. fascicularis* has prompted the Department of Wildlife and National Parks (DWNP) to implement several population control measures including translocations and culling [5,30]. Conflict *M. fascicularis* mainly male individuals, have been translocated from conflict urban areas to other locations with a less inhabited human population [5]. Therefore, in this study, by investigating the SRY and TSPY segments, information on whether Y-DNA of *M. fascicularis* consists of continuously related or segregated haplotypes would provide information on whether the translocated individuals posed any risks of contamination to the existing population. Furthermore, taking into account the historical zoogeography and phylogeography of SEA, investigation into the male-mediated gene flow would also provide insight into the hypothesis on the existence of historical or current barrier as well as evidence of recent gene flow among the *M. fascicularis* populations in SEA, particularly in Malaysia.

Results

Haplotype Mapping, Sequence Characteristics, and Genetic Diversity

A total of 654- and 1,370-bp of sequence lengths were obtained from the sex-determining region (SRY) and the testis-specific protein (TSPY) segments, respectively, from all 164 male *M. fascicularis* samples. Since both segments are closely linked on the Y-chromosome, the datasets from 287 sequences were combined to produce an alignment of 2,032-bp (including indels). All the sequences in this study were registered with the National Center for Biotechnology Information (NCBI) and were given accession numbers, as shown in Additional file 1: Table S1.

Considering all the *M. fascicularis* sequences (N = 232), 20 haplotypes were observed. Out of the 20 haplotypes, the Malaysian *M. fascicularis* were represented by 18 haplotypes in which 16 were unique to Peninsular Malaysia (Additional file 1: Table S1; Fig. 2a). Peninsular Malaysia constitutes 17 haplotypes where Haplotype 1 is the most dominant and can be found in all the states and major islands (except on Tioman Island). Interestingly, Haplotype 1 is also shared with all nine *M. fascicularis* sequences from the southern Thai Peninsular (south of Surat Thani-Krabi depression), four sequences from Sumatra (Fig. 2b), and the introduced population to Mauritius (N= 10). The Indochinese *M. fascicularis* sequences from Vietnam, Cambodia, Thailand (north of Surat Thani-Krabi depression) formed a single haplotype, Haplotype 20 (Additional file 2: Table S2; Fig. 2b) and were uniquely shared with *M. mulatta* from Burma and South China. The Malaysian Borneo sequences (Sarawak, Labuan Island, and Sabah) formed a single haplotype (Haplotype 18) shared with sequences from Kalimantan and the Philippines including two sequences from southern Sumatra while Java was represented by Haplotype 19.

Based on the haplotype mapping, the *M. fascicularis* sequences were assigned into three geographical groups (except for sequences from Mauritius) for further sequence characterization: (1) Indochinese (Haplotype 20); (2) Continental Sundaic consisting of sequences from the Thai peninsula, Peninsular Malaysia, and Sumatra (Haplotype 1 – 17), and; (3) Insular Sundaic (Haplotype 18 – 19). Four indels were observed in *M. fascicularis*, distinguishing between the Indochinese and the Sundaic (continental and insular) haplotypes where two events of; (1) three-bp indels (ACA) at nucleotide position (np) 596–598 and (2) one-bp thymine (T) indel at np 974 occurred (Table 1). Furthermore, within the Sundaic lineage, a single transversion mutation (adenine to thymine) at np 159 was observed on the SRY segment which distinguished the insular Sundaic from the continental Sundaic haplotypes.

Table 1

The observed haplotypes and variable sites of the SRY (N= 5) and TSPY (N= 16) segments of *M. fascicularis*. A single nucleotide polymorphism (transversion mutation at np 159) was observed (in red font) differentiated the insular Sundaic group from the continental Sundaic group while four indels (np 596-598 and 974; in blue font) distinguished the Sundaic group from the Indochinese *M. fascicularis*

Geographical Groups	Hap. No.	N	SRY					TSPY															
			1	2	5	5	5	6	7	8	9	9	0	1	3	3	4	6	8	9	9	0	0
Continental Sundaic	Hap1	93	A	A	-	-	-	T	C	A	T	A	A	A	A	A	A	A	A	T	G	G	G
	Hap2	1	.	.	-	-	-	G
	Hap3	25	.	.	-	-	-	G
	Hap4	8	.	.	-	-	-	G
	Hap5	12	.	.	-	-	-	C	.
	Hap6	4	.	.	-	-	-	G	G
	Hap7	1	.	.	-	-	-	G	G	.	.	T	.
	Hap8	1	.	.	-	-	-	G	T	.
	Hap9	2	.	.	-	-	-	T	.
	Hap10	6	.	.	-	-	-	G
	Hap11	1	.	.	-	-	-	G	.	.	T	.
	Hap12	1	.	.	-	-	-	A
	Hap13	1	.	.	-	-	-	C	.	.	.
	Hap14	1	.	C	-	-	-	G	G
	Hap15	2	.	.	-	-	-	G	.	T	C
	Hap16	1	.	.	-	-	-	T	G
	Hap17	1	.	.	-	-	-	.	.	C
Insular Sundaic	Hap18	26	T	.	-	-	-	
	Hap19	1	T	.	-	-	-	.	.	.	G	
Indochinese	Hap20	34	T	.	A	C	A	.	.	.	-	.	.	C	C	.	G	

Table 2 summarized the sequence characterization, genetic diversity indices, genetic structure, and differentiation among the geographical groups of *M. fascicularis*. The continental Sundaic is the most variable group displaying the highest number of variable sites (VS) and parsimony-informative sites (PIS). Estimation of the Y-DNA nucleotide diversity (π) among the Sundaic lineage revealed that the continental Sundaic haplotypes had a significantly higher π (4.4×10^{-4}) compared to the insular Sundaic (0.4×10^{-4}). The genetic diversity for the Indochinese lineage was incalculable since only a single haplotype was observed. The pairwise fixation index (F_{ST}) calculated among the groups revealed that the Indochinese *M. fascicularis* is highly divergent from both the Sundaic forms (F_{ST} of 0.90 and 0.99 to the continental and insular forms, respectively) while within the Sundaic lineage, the insular and continental forms are moderately differentiated from each other (F_{ST} of 0.69). The genetic distances among the haplotypes range from 0.05 – 0.35% (data not shown). Among the geographical groups, the genetic distance between continental Sundaic and insular Sundaic is at 0.15%, while the genetic distances between Indochinese with the continental Sundaic and insular Sundaic is at 0.27% and 0.17%, respectively.

Table 2

Sequence characterization (as calculated using MEGA7), standard molecular diversity indices (calculated using DnaSP) as well as pairwise F_{ST} and genetic distances (%) among the geographical groups of *M. fascicularis*. CS= conserved sites; VS= variable sites; PIS= parsimony-informative sites; NHap= number of haplotypes; Hd= haplotype diversity; π = nucleotide diversity

Geographical Groups	N	Sequence Characterization			Diversity Indices			Pairwise F_{ST} (below diagonal) and genetic distances (%; above diagonal)		
		CS	VS	PIS	NHap	Hd	π (10^{-4})	1	2	3
1 Indochinese	34	2028	0	0	1	0	0		0.27	0.17
2 Continental Sundaic	171	2014	12	7	17	0.61	4.40	0.90		0.15
3 Insular Sundaic	27	2025	1	0	2	0.07	0.40	0.99	0.69	

Phylogenetic Relationships and Estimation of Divergences

The combined SRY and TSPY phylogenetic relationships constructed by using the maximum likelihood (ML) (-lnL = 3809.0846) and Bayesian inference (BI) methods produced similar topologies and thus were represented by the BI tree (Fig. 3). In general, the tree grouped the *Macaca* into their respective species groupings consistent with the four major lineages of *Macaca*: (1) *sylvanus-silenus*, (2) *sinica*, (3) *arctoides*, and (4) *fascicularis*.

Within the *fascicularis* species group, the tree formed two distinct clades: (1) a clade consisting of *M. cyclopis*, *M. fuscata*, *M. mulatta*, and the Indochinese *M. fascicularis*, and (2) *M. fascicularis* haplotypes of Sundaic origin, with high support value (99% and 100% for ML and BI, respectively). Furthermore, the Sundaic *M. fascicularis* lineage further bifurcated into the continental and insular forms but with low support values (below 50%). Similarly, the median-joining network (MJN) (Fig. 4) produced similar groupings, as observed in the phylogenetic tree. Moreover, MJN showed that the continental Sundaic haplotypes were all derived from the insular form, particularly from Haplotype 18. Furthermore, all the other haplotypes within the continental Sundaic haplotypes were derived from the dominant Haplotype 1.

The divergence date estimations calculated by calibrating the BI using the proposed date for the last common ancestor (LCA) of *Macaca* at 5.5 million years ago (mya) are shown in Fig. 3. The *fascicularis* species group (*cyclopis*, *fascicularis*, *fuscata*, and *mulatta*) shared the LCA at around 2.35 mya. Next, the bifurcation between the continental and insular Sundaic forms was estimated to have occurred at ~2.00 mya. After that, the continental form shared the LCA at around 1.65 mya while the insular form shared the LCA at around 0.71 mya.

Discussion

By analyzing 164 male samples from 62 localities throughout Malaysia (including five major islands), this work represents the most comprehensive effort to infer male-mediated gene flow of *M. fascicularis* in Malaysia based on the TSPY and SRY segments of the Y-DNA gene. Combining this dataset with the sequences of *M. fascicularis* from the other regional populations, a complete and broad analysis of the Y-DNA gene flow were able to be estimated, covering the entire range of the species in SEA. In general, the combined segments of the SRY and TSPY in this study were able to provide several paternal insights into the gene flow, phylogenetic relationships at the population level as well as divergence time estimation of *M. fascicularis*.

In this study, the phylogenetic relationship of *Macaca* based on the Y-DNA supported the four major species groupings congruent with the morphological classification by Delson [31]: *sylvanus*, *silenus*, *sinica*, and *fascicularis* species groups. Although supported with low bootstrap values (except for *fascicularis* species group which is supported by high bootstrap values of 99 and 100% for BI and ML, respectively), all *Macaca* species were clustered into their respective species groups. Similarly, previous nuclear DNA studies also supported the four major groupings [9,16,32-34].

Within the *fascicularis* species group, two major clades were observed: (1) a species complex group comprising of *M. cyclopis*, *M. fuscata*, *M. mulatta*, and the Indochinese *M. fascicularis* and (2) *M. fascicularis* haplotypes of Sundaic origin. Both clades shared an LCA at ~2.35 mya, which is in proximity to the previous estimation of 2.30 mya by Tosi, Morales, and Melnick [9]. This estimation coincides with the multiple major climate cooling events during Late Pliocene and early Pleistocene occurring between 2.8 and 2.4 mya

[35-36] that influenced the paleoenvironments of SEA [24-25]. The cold and dry climate during the glacials gave rise to savannah-like vegetation or semi-open woodland [25,27,29,37], causing restriction of dispersal for primates, thus could lead to the separation between both clades.

Hybridization of *M. mulatta* Y-DNA into *M. fascicularis* in the Indochinese Region

The haplotype sharing between *M. mulatta* (N = 12) and *M. fascicularis* (N = 34) represents the male-mediated gene flow of *M. mulatta* into *M. fascicularis* populations from the Indochinese region. Several authors have previously reported on the possible occurrence of the hybridization events between the Indochinese *M. fascicularis* and *M. mulatta* [9,15-16,23,38-40]. The extent of the contamination of *M. mulatta* Y-DNA into the Indochinese *M. fascicularis* population is not only restricted to the contact zones between both species but possibly more widespread across the Indochinese region [23,38-39].

In this study, we confirmed the absence of the male-mediated gene flow of *M. mulatta* into *M. fascicularis* from the Sundaic region, which is restricted at the Surat Thani-Krabi depression region, as shown in Fig. 2b. This is similar to the separation observed between the southern and northern pig-tailed macaques [41]. It is possible that the Surat Thani-Krabi depression played a significant role as a barrier that limits the downwards dispersal of the contaminated Indochinese *M. fascicularis* into the Sundaic lineage. Thus, based on the current finding, we defined the Sundaic *M. fascicularis* as the original carrier of Y-DNA in *M. fascicularis*.

The region of Isthmus of Kra and Kangar-Patani marks the biogeographical barrier that separated the Indochinese and Sundaic zoogeographical regions [42]. These regions display the shift from mixed deciduous forest to the north to wet seasonal rainforest type to the south [43]. Situated along the modern Thailand–Malaysian border (see Fig. 2b), these biogeographical barriers has been documented to limit the gene flow between the Indochinese *M. fascicularis* from their Sundaic conspecifics [15,23,39,44] as well as in other vertebrate faunal species [43,45-50]. Therefore, to measure the extent of the *M. mulatta* male hybridization into *M. fascicularis*, further research to screen *M. fascicularis* from this area are needed.

Sundaic *M. fascicularis*: Continental and Insular Forms

The Sundaic *M. fascicularis* lineage, which is the original carrier of Y-DNA in *M. fascicularis*, was further separated into continental and insular forms and shared the LCA at about 2.00 mya. The continental form comprised of haplotypes from southern Thai Peninsular (south of Surat Thani-Krabi depression), Peninsular Malaysia, and Sumatra, including the introduced Mauritian population, while the insular form consisted of haplotypes from southern Sumatra, Java, Borneo, and the Philippines. Likewise, previous molecular studies have identified both forms using mtDNA and nuclear gene [15,23,51-56]. Furthermore, the MJN suggested that the insular form was the predecessors of the continental form (Fig. 4). This result would be consistent with the hypothesis that the ancestors of *M. fascicularis* evolved in insular SEA [31,57-58]. Wherever and whenever this species originated, the earliest *M. fascicularis*-like fossil was discovered in Java, suggesting that they have already inhabited the Sundaic region as early as 0.9 – 1.0 mya [58].

After diverging out from the insular form, the continental form diverged into 17 haplotypes, which all shared the LCA at around 1.65 mya (Fig. 3). The dominant Haplotype 1 characterized this form, which is shared with 103 individuals and is distributed as far north as Wat Suwankhuha, of the Thai Peninsular, downwards to entire Peninsular Malaysia until Sumatra of Indonesia. According to the MJN (Fig. 4), all the other continental haplotypes were derived from this dominant haplotype. Moreover, within Peninsular Malaysia, the distribution of haplotypes (Fig. 2a) further showed the segregation of several unique haplotypes that can be found only at the north-western region (Haplotype 4, 6 – 8, 11 – 14) as well as the east coast region (Haplotype 15 – 17). In contrast, the central-southern regions are relatively connected with no unique haplotype defining the region.

Compared to the 17 observed haplotypes within the continental form, only two haplotypes were observed within the insular form, which shared the LCA at around 0.71 mya (Fig. 3). *M. fascicularis* from southern Sumatra, Borneo, and the Sibuyan Island of northern Philippines were all represented by a single dominant insular haplotype, Haplotype 18. The island of Java, on the other hand, was represented by Haplotype 19. The small number of haplotypes observed shows that the insular form consists of significantly low

variability and intact Y-DNA. The highly dispersal nature of males would move and homogenize Y-DNA across populations [9] and therefore contribute to this highly intact Y-DNA across the insular region.

Prolong and recurring connectivity that existed in the past could have facilitated the dispersal of males between the insular regions (Sumatra, Java, Borneo, and the Philippines). Several episodes of sea-level fluctuations in Sundaland during the Pliocene up until recently during the Holocene around seven thousand years ago (kya) would permit for such connectivity with some as low as -120 m below the present level [25,59]. In addition to the connectivity, a possible ancient bottleneck could explain for the intact Y-DNA in the insular region. The ancestors of the insular form could have been forced into a common refugium postulated to have existed in northern Borneo [24-25,29,60-61]. Over time, only a single dominant haplotype could have persevered. Alternatively, overhunting and exploitation by early human could also severely have reduced the effective Y-DNA gene pool. Remains of *M. fascicularis*-like species were discovered in cave settlements in Java (~0.9 – 1.0 mya) and Borneo (~30 – 40 kya) [58] giving evidence to early exploitations by human. Nevertheless, the low number of samples and available sequences (N = 27) could hinder the possibility of detecting other haplotypes within the insular form. Therefore, to further measure the diversity of the insular form, further research utilizing more samples should be conducted on the insular form.

Uniquely, *M. fascicularis* from Sumatra exhibited both the continental and insular Y-DNA haplotypes (Fig. 2b; Additional file 2: Table S2). A secondary contact hypothesis could account for this condition in Sumatra with two possible explanation. The first explanation offered by Tosi and Coke [62] hypothesized that Sumatra was already populated by the insular Sundaic form, prior to the immigrating males of the continental form from the Malay Peninsula. Due to the narrow and shallow waterways on the Sunda Shelf separating the Malay Peninsula and Sumatra [25,59], a recent dispersal of the dominant continental haplotype from the Malay Peninsula is postulated as being in the process of colonizing the native insular form.

In contrast, a second explanation similar to the first secondary contact hypothesis takes into account the biogeographical history of both regions. Several authors have postulated the formation of several refugia during the Pleistocene, whereas the Malay Peninsula was surrounded by a belt of savannah-like vegetation [24-26,63]. This implies that dispersal of the continental form occurred from a postulated refugium in Sumatra into the Malay Peninsula and the surrounding areas. Dispersal of the insular form from a refugium in western Java into Sumatra, which was already populated by the continental Sundaic haplotypes could explain for the presence of both forms in Sumatra.

Conclusions

Findings in this study could also assist the authorities particularly in Malaysia on the population management of conflict *M. fascicularis*. Prior information on the genetic structure is imperative before any measures of population control such as translocation programs or culling are to be conducted [64-66]. Thus, the haplotype mapping in Fig. 2a can be used by the managing authority to correctly manage the local populations without posing risks of contamination to the existing population or accidentally wiping out a unique haplotype due to culling.

Overall, Fig. 2b summarizes the Y-DNA mapping of *M. fascicularis* according to the lineage and forms observed. Evidently, *M. fascicularis* are distributed accordingly to the zoogeographical segregation of SEA. The single nucleotide polymorphism observed at np 159 which distinguishes between the continental and insular forms are important from the perspective of biomedical research. The identification of the genetic background of non-human primate model used in experimental research is important since variations could influence the results and repeatability of experiments [67-68]. Thus, this finding could assist in the identification of the origin of the captive *M. fascicularis* used in biomedical research, at least paternally.

Finally, findings of this study suggest that the separation between the continental and insular forms warrants for the treatment as separate management units (MUs). MUs are lower ranked conservation units which recognize populations with significant divergence at nuclear or mtDNA and represent populations connected by such low levels of gene flow [65]. The status of the Indochinese lineage, on the other hand, should be handled with caution and until the level of genetic hybridization of Indochinese *M. mulatta* into *M. fascicularis* populations are thoroughly assessed, this lineage should be managed separately from the Sundaic lineage.

Methods

Sample Collection and GenBank Sequences

Blood samples were collected from wild free-ranging conflict *M. fascicularis* by authorized personnel and veterinarians of DWNP (for samples from Peninsular Malaysia), Sabah Wildlife Department (for samples from Sabah), and Forest Department of Sarawak (for samples from Sarawak). A total of 164 samples of adult male long-tailed macaques from 62 localities in Peninsular Malaysia, Sarawak, and Sabah were collected and used in this study (Additional file 1: Table S1). Fig. 1a and b summarizes the localities of the samples according to the states and major islands in Malaysia.

Sixty-eight sequences of *M. fascicularis* from other regional populations were also included in the dataset to make up a total of 232 sequences (Fig. 1c; Additional file 2: Table S2) representing the most comprehensive sampling coverage of the species distribution range. In total, 287 sequences were used for analyses comprising of 20 *Macaca* species and two outgroup species (*Papio hamadryas* and *Mandrillus sphinx*). Both outgroup species were selected as representatives from the major branches within the Family Cercopithecidae.

DNA Extraction, PCR Amplification, and DNA Sequencing

Total genomic DNA was extracted from blood samples using the QIAamp DNeasy® Blood and Tissue Kit using the protocol provided by the manufacturer (Qiagen, Germany). Two pairs of oligonucleotides were used to amplify the and the TSPY segments of the Y-DNA as designed by Rovie-Ryan et al. [21]. In Old World monkeys the SRY segment is found on the short arm of the Y-DNA [69] while the TSPY segment is typically located on the long arm, close to the centromere region in the Y-DNA [70-71].

PCR amplifications were conducted in 20 µl reactions using a GeneAmp® PCR System 9700 (Applied Biosystems, USA). Each PCR reaction consists of 1.0 µl of DNA template (~15 – 20 ng), PCR mixtures containing 4X Green GoTaq® Flexi Buffer (Promega, USA), 0.875 mM of MgCl₂ (25 mM), 0.1 mM of each dNTPs (10 mM), 0.1 mM of each primer (10 mM), 1 unit of *Taq* Polymerase (5 unit/µl), and later added with ddH₂O to a total of 20 µl of total reaction mixtures. The cycling profile for the amplification was as follows: a preliminary denaturation at 98°C for 2 min followed by 45 cycles of 95°C for 30 sec, 55°C for 60 sec and 72°C for 60 sec. This was followed by a final extension period of 72°C for 5 min before the samples were cooled to 10°C. Cycle sequencing was done on an ABI PRISM®377 DNA Sequencer by a sequencing service provider (1st Base Laboratories Sdn. Bhd., Malaysia).

Sequence Characterizations and Genetic Diversity Indices

Multiple sequence alignments were done by using the program Geneious ver5.6 [72]. Prior to further sequence analysis, the SRY and TSPY segments were combined as both segments are closely linked on the Y-chromosome and the partition homogeneity tests conducted in PAUP ver4 [73] did not find significant differences in their evolutionary signal ($P = 1.0$) as also previously reported [9,16].

Sequence characterizations including conserved sites (CS), VS, and PIS were examined by using MEGA ver6 [74] while DnaSP ver5 [75] were used to calculate the standard genetic diversity indices including the number of haplotypes (NHap), haplotype diversity (Hd), and π [76]. Genetic distances were calculated on MEGA ver6 using the Kimura 2-parameter model [77]. To evaluate the amount of population genetic structure pairwise F_{ST} values were calculated using DnaSP ver5.

Phylogenetic Analysis and Divergence Time Estimates

To infer phylogenetic relationships, the haplotypes data were used to construct phylogenetic trees using the ML, BI and the MJN methods. The best substitution model to run the ML tree and BI was Kimura 2-parameter model with discrete gamma distribution (K2P + G) as determined using MEGA ver6.1. To assess the robustness of the ML tree, bootstrapping [78] with 2,000 replicates were conducted. In BEAST package ver2.5 [79], BI was analyzed using two independent runs with 10 million of MCMC chain length each using strict clock settings, Yule model prior [80], and sub-sampled at every 1,000 generations. To set the substitution model to K2P + G, HKY85 model [81] was selected with all the base frequencies set to equal and all rate parameters were fixed to 1.0. The convergence of all parameters was

assessed using TRACER ver1.6 [82] and both independent runs were then combined using the software LogCombiner ver2.5 available within the BEAST ver2.0 package. A consensus tree was later created from the combined tree files after a burn-in of 10% using the TreeAnnotator ver2.5 also available within the BEAST ver2.0 package. Finally, the MJN was constructed using the Network ver4.6 [83].

To estimate the divergence dates of several important events in the evolutionary and dispersal history of *M. fascicularis* in SEA, the BI tree was recalibrated using the same BI analysis parameters as described above. Using the proposed divergence date of 5.5 mya for the LCA of *Macaca* [9,31,84], the priors were set to normal distribution with the mean set to 5.5 and sigma at 0.5.

Abbreviations

BI: Bayesian inference; CS: Conserved site; DWNP: Department of Wildlife and National Parks; kya: Thousand years ago; LCA: Last common ancestor; MJN: Median-joining network; ML: Maximum likelihood; mtDNA: Mitochondrial DNA; MU: Management unit; mya: Million years ago; NCBI: National Center for Biotechnology Information; PIS: Parsimony-informative site; SEA: Southeast Asia; SRY: Sex-determining region; TSPY: Testis-specific protein; VS: Variable site

Declarations

Ethics approval and consent to participate

Sampling at Peninsular Malaysia was permitted by the Research Committee of DWNP, at Sarawak by Forest Department of Sarawak (Permit No: 37/2011), and at Sabah by the Sabah Biodiversity Council (Permit No: JKM/MBS.1000-2/2(103)). The sampling protocol was approved by UC DAVIS Institutional Animal Care and Use Committee (protocol number: 16048).

Consent for publication

Not applicable.

Availability of data and materials

Sequence data produced in this study are available in the NCBI (<https://www.ncbi.nlm.nih.gov/>) with the Accession No.: KC572634-KC572687; KJ690361-KJ690376; KJ733018-KJ733278.

Competing interests

The authors declare that they have no competing interests.

Funding

This project was funded by the DWNP for laboratory equipment and consumables, and partly by the Proboscis Genome Research grant for travel and accommodation, awarded to M. T. Abdullah and colleagues. Funding for sampling work for ZMW coded samples was supported in part by the USAID Emerging Pandemic Threat Program – PREDICT Project and by the Skoll Foundation and Google Incorporated through Global Viral.

Authors' contributions

This study was conceived by JJRR and MTA. Sample collection, lab analysis and data analysis were performed by JJRR. JJRR, FAAK and MTA wrote the paper. All authors read and approved the final manuscript.

Acknowledgment

We acknowledge the following local authorities for the permission collect samples: Department of Wildlife and National Parks (DWNP), Forest Department of Sarawak, and Sabah Wildlife Department. We are thankfully to the Outbreak Response Team of DWNP, UNIMAS Primate Team, and EcoHealth Alliance for assistance during sampling.

References

1. Fooden J, Albrecht GH. Latitudinal and insular variation of skull size in crab-eating Am J Phys Anthropol. 1993;92:521-38.
2. Fooden J. Systematic review of Southeast Asian long-tail macaques, *Macaca fascicularis* (Raffles, 1821). Fieldiana Zool. 1995;81:1-206.
3. Eudey AA. The crab-eating macaque (*Macaca fascicularis*): Widespread and rapidly Primate Conser. 2008;23:129-32.
4. Osman A study on composition and density of long-tailed macaque (*Macaca fascicularis*) in Kuala Lumpur, 1998. JWP. 1998;16:80-4.
5. Pelan pengurusan kera (*Macaca fascicularis*) bermasalah di Semenanjung Malaysia. Department of Wildlife and Parks. 2006. https://www.wildlife.gov.my/images/stories/penerbitan/lain_lain/PelanPengurusanKeraBermasalah.pdf. Accessed 11 Oct 2016.
6. Payne J, Francis CM. A field guide to the mammals of Borneo. Kota Kinabalu: Sabah Society; 2005.
7. Pusey AE, Packer C. Dispersal and philopatry. In: Smuts BB, Cheney DL, Seyfarth RM, Wrangham RW, Struhsaker TT, editors. Primate societies. Chicago: University of Chicago Press; 1987. p. 250-66.
8. Melnick DJ, Hoelzer GA. Differences in male and female macaque dispersal lead to contrasting distributions of nuclear and mitochondrial DNA variation. Int J Primatol. 1992;13:379-93.
9. Tosi AJ, Morales JC, Melnick DJ. Paternal, maternal, and biparental molecular markers provide unique windows onto the evolutionary history of macaque monkeys. Evolution. 2003;57:1419-35.
10. Avise JC. Phylogeography: Retrospect and prospect. Biogeogr. 2009;36:3–15.
11. Jobling MA, Tyler-Smith C. Fathers and sons: The Y chromosome and human evolution. Trends Genet. 1995;11:449-56.
12. Keane B, Dittus WPJ, Melnick DJ. Paternity assessment in wild groups of toque macaques *Macaca sinica* at Polonnaruwa, Sri Lanka using molecular markers. Mol Ecol. 1997;6:267-82.
13. Disotell TR. Human evolution: Sex-specific contributions to genome variation. Curr Biol. 1999;9:R29-31.
14. Chan YC, Roos C, Inoue-Murayama M, Inoue E, Shih CC, Vigilant L. A comparative analysis of Y-chromosome and mtDNA phylogenies of the *Hylobates* BMC Evol Biol. 2012;12:150.
15. Tosi AJ, Morales JC, Melnick DJ. Y-chromosome and mitochondrial markers in *Macaca fascicularis* indicate introgression with Indochinese *mulatta* and a biogeographic barrier in the Isthmus of Kra. Int J Primatol. 2002;23:161-77.
16. Tosi AJ, Morales JC, Melnick DJ. Comparison of Y-chromosome and mtDNA phylogenies leads to unique inferences of macaque evolutionary history. Mol Phylogenet Evol. 2000;17:133-44.
17. Tosi AJ, Buzzard PJ, Morales JC, Melnick DJ. Y-chromosome data and tribal affiliations of *Allenopithecus* and *Miopithecus*. Int J Primatol. 2002;23:1287-99.
18. Tosi AJ, Disotell TR, Morales JC, Melnick DJ. Cercopithecine Y-chromosome data provide a test of competing morphological evolutionary hypotheses. Mol Phylogenet Evol. 2003;27:510-21.
19. Tosi AJ, Melnick DJ, Disotell TR. Sex chromosome phylogenetics indicate a single transition to terrestriality in the guenons (tribe Cercopithecini). J Hum Evol. 2004;46:223-37.
20. Tosi AJ, Detwiler KM, Disotell TR. Y-chromosomal markers suitable for noninvasive studies of guenon hybridization. Int J Primatol. 2005;26:685-96.
21. Rovie-Ryan JJ, Abdullah MT, Sitam FT, Abidin ZZ, Tan SG. Y-chromosomal gene flow of *Macaca fascicularis* (Cercopithecidae) between the insular and mainland peninsula of Penang state, Malaysia. JoSTT 2013;9:113-26.
22. Bunlungsup S, Imai H, Hamada Y, Gumert MD, San AM, Malaivijitnond S. Morphological characteristics and genetic diversity of Burmese long-tailed Macaques (*Macaca fascicularis aurea*). Am J Primatol. 2016;78:441-55.

23. Bunlungsup S, Imai H, Hamada Y, Matsudaira K, Malaivijitnond S. Mitochondrial DNA and two Y-chromosome genes of common long-tailed macaques (*Macaca fascicularis fascicularis*) throughout Thailand and vicinity. *Am J Primatol.* 2017;79:e22596.
24. Abegg C, Thierry B. Macaque evolution and dispersal in insular south-east Asia. *Biol J Linn.* 2002;75:555-76.
25. Bird MI, Taylor D, Hunt C. Palaeoenvironments of insular Southeast Asia during the Last Glacial Period: A savanna corridor in Sundaland? *Quat Sci Rev.* 2005;24:2228-2242.
26. Harrison T, Krigbaum J, Manser J. Primate biogeography and ecology on the Sunda shelf Islands: a paleontological and zooarchaeological perspective. In: Lehman SM & Fleagle JG, editors. *Primate biogeography: Progress and prospects.* New York: Springer Science; 2006. p. 331-72.
27. Cannon CH, Morley RJ, Bush ABG. The current refugial rainforests of Sundaland are unrepresentative of their biogeographic past and highly vulnerable to disturbance. 2009;106:11188-93.
28. Arora N, Nater A, van Shaik CP, Willems EP, van Noordwijk MA, Goossens B, Morf N, Bastian M, Knott C, Morrogh-Bernard H, Kuze N, Kanamori T, Pamungkas J, Perwitasari-Farajallah D, Verschoor E, Warren K, Krutzen M. Effects of Pleistocene glaciations and rivers on the populations structure of Bornean orangutans (*Pongo pygmaeus*). 2010;107:21376-81.
29. Wurster CM, Bird MI, Bull ID, Creed F, Bryant C, Dungait JAJ, Paz V. Forest contraction in north equatorial Southeast Asia during the Last Glacial Period. 2010;107:15508-11.
30. Karuppannan KV, Saaban S, Firdaus Ariff AR, Mustapa AR. Non-surgical castration in controlling long tailed macaques (*Macaca fascicularis*) population by Department of Wildlife and National Parks (DWNP) Peninsular Malaysia. 2013;4:33-6.
31. Delson E. Fossil macaques, phyletic relationships and a scenario of deployment. In: Lindburg DG, editor. *The macaques: Studies in ecology, behavior, and evolution.* New York: Van Nostrand Reinhold Co; 1980. p. 10–30.
32. Deinard A, Smith DG. Phylogenetic relationships among the macaques: Evidence from the nuclear locus *NRAMP1*. *J Hum Evol.* 2001;41:45–59.
33. Li J, Han K, Xing J, Kim HS, Rogers J, Ryder OA, Disotell T, Yue B, Batzer MA. Phylogeny of macaques (Cercopithecidae: *Macaca*) based on *Alu Gene*. 2009;448:242-9.
34. Fan Z, Zhou A, Osada A, Yu J, Jiang J, Li P, Du L, Niu L, Deng J, Xu H, Xing J, Yue B, Li J. Ancient hybridization and admixture in macaques (genus *Macaca*) inferred from whole genome sequences. *Mol Phylogenet Evol.* 2018;127:376-86.
35. Gibbard PL, Head MJ, Walker MJ. Formal ratification of the Quaternary System/Period and the Pleistocene Series/Epoch with a base at 2.58 Ma. *J Quat Sci.* 2010;25:96-102.
36. Gibbard PL, Head MJ. The newly-ratified definition of the Quaternary System/Period and redefinition of the Pleistocene Series/Epoch, and comparison of proposals advanced prior to formal ratification. 2010;33:152-8.
37. Cranbrook E, Piper PJ. Paleontology to policy: The Quaternary history of Southeast Asian tapirs (Tapiridae) in relation to large mammal species turnover, with a proposal for conservation of Malayan tapir by reintroduction to Borneo. *Integr Zool.* 2013;8:95-
38. Osada N, Uno Y, Mineta K, Kameoka Y, Takahashi I, Terao K. Ancient genome-wide admixture extends beyond the current hybrid zone between *Macaca fascicularis* and *mulatta*. *Mol Ecol.* 2010;19:2884-95.
39. Bunlungsup S, Kanthaswamy S, Oldt RF, Smith DG, Houghton P, Hamada Y, Malaivijitnond S. Genetic analysis of samples from wild populations opens new perspectives on hybridization between long-tailed (*Macaca fascicularis*) and rhesus macaques (*Macaca mulatta*). *Am J Primatol.* 2017;79:e22726.
40. Yan G, Zhang G, Fang X, Zhang Y, Li C, Ling F, Cooper DN, Li Q, Li Y, van Gool AJ, Du H, Chen J, Chen R, Zhang P, Huang Z, Thompson JR, 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 PD, Li B, Liu X, Ball EV, An N, Huang Q, Zhang Y, Fan W, Zhang X, Li Y, Wang W, Katze MG, Su B, Nielsen R, Yang H, Wang J, Wang X, Wang J. Genome sequencing and comparison of two nonhuman primate animal models, the cynomolgus and Chinese rhesus macaques. *Nat Biotechnol.* 2011;29:1019-23.
41. Malaivijitnond S, Arsaithamkul V, Tanaka H, Pomchote P, Jaroenporn S, Suryobroto B, Hamada Y. Boundary zone between northern and southern pig-tailed macaques and their morphological Primates 2012;53:377-89.
42. Wallace AR. *The geographical distribution of animals.* London: Macmillan; 1876.
43. Woodruff DS. Neogene marine transgressions, palaeogeography and biogeographic transitions on the Thai–Malay Peninsula. *J Biogeogr.* 2003;30:551-67.
44. Doxiadis GGM, de Groot N, de Groot NG, Rotmans G, de Vos-Rouweler AJM, Bontrop RE. Extensive *DRB* region diversity in cynomolgus macaques: Recombination as a driving force. *Immunogenetics* 2010;62:137-47.

45. Inger RF, Voris HK. The biogeographical relations of the frogs and snakes of Sundaland. *J Biogeogr.* 2001;28:863-91.
46. Abdullah MT. Biogeography and variation of *Cynopterus brachyotis* in Southeast Asia. D. Phil. Thesis. Queensland: University of Queensland; 2003.
47. Hughes JB, Round PD, Woodruff DS. The Indochinese–Sundaic faunal transition at the Isthmus of Kra: An analysis of resident forest bird species distributions. *J Biogeogr.* 2003;30:569-80.
48. Round PD, Hughes JB, Woodruff DS. Latitudinal range limits of resident forest birds in Thailand and the Indochinese–Sundaic zoogeographic transition. *Nat Hist Bull Siam Soc.* 2003;51:69-96.
49. Luo SJ, Kim JH, Johnson WE, van der Walt J, Martenson J, Yuhki N, Miquelle DG, Uphyrkina O, Goodrich JM, Quigley HB, Tilson R, Brady G, Martelli P, Subramaniam V, McDougal C, Hean S, Huang SQ, Pan W, Karanth UK, Sunquist M, Smith JLD, O'Brien SJ. Phylogeography and genetic ancestry of tigers (*Panthera tigris*). *PLoS Biol.* 2004;2:2275-93.
50. Woodruff DS, Turner LM. The Indochinese-Sundaic zoogeographic transition: A description and analysis of terrestrial mammal species distributions. *J Biogeogr.* 2009;36:803-21.
51. Smith DG, McDonough JW, George DA. Mitochondrial DNA variation within and among regional populations of longtail macaques (*Macaca fascicularis*) in relation to other species of the *fascicularis* group of macaques. *Am J Primatol.* 2007;68:182-98.
52. Blancher A, Bonhomme M, Crouau-Roy B, Terao K, Kitano T, Saitou N. Mitochondrial DNA sequence phylogeny of 4 populations of the widely distributed cynomolgus macaque (*Macaca fascicularis fascicularis*). *J Hered.* 2008;99:254-64.
53. Shiina T, Tanaka K, Katsuyama Y, Otabe K, Sakamoto K, Kurata M, Nomura M, Yamanaka H, Nakagawa H, Inoko H, Ota M. Mitochondrial DNA diversity among three subpopulations of cynomolgus macaques (*Macaca fascicularis*) originating from the Indochinese region. *Exp Anim.* 2010;59:567-78.
54. Abdul-Latiff MAB, Ruslin F, Faiq H, Hairul MS, Rovie-Ryan JJ, Abdul-Patah P, Yaakop S, Md-Zain BM. Continental monophyly and molecular divergence of peninsular Malaysia's *Macaca fascicularis fascicularis*. *BioMed Res Int.* 2014;897682.
55. Abdul-Latiff MAB, Ruslin F, Fui VV, Abu MH, Rovie-Ryan JJ, Abdul-Patah P, Lakim M, Roos C, Yaakop S, Md-Zain BM. Phylogenetic relationships of Malaysia's long-tailed macaques, *Macaca fascicularis*, based on cytochrome *b* ZooKeys. 2014;407:121-39.
56. Liedigk R, Kolleck J, Böker KO, Meijaard E, Md-Zain BM, Abdul-Latiff MAB, Ampeng A, Lakim M, Abdul-Patah, P, Tosi AJ, Brameier M, Zinner D, Roos C. Mitogenomic phylogeny of the common long-tailed macaques (*Macaca fascicularis fascicularis*). *BMC Genomics.* 2015;16:222.
57. Fa JE. The genus *Macaca*: A review of taxonomy and evolution. *Mammal Rev.* 1989;19:45-81.
58. Fooden J. Comparative review of *fascicularis*-group species of macaques (Primates: *Macaca*). *Fieldiana Zool.* 2006;107:1-43.
59. Voris HK. Maps of Pleistocene sea levels in Southeast Asia: Shorelines, river systems and time durations. *J Biogeogr.* 2000;27:1153-67.
60. Meijaard E, Groves CP. Biogeography and phylogeography of *Presbytis*. *Primate Rep.* 2004;68:71-90.
61. Stimpson CM. Local scale, proxy evidence for the presence of closed canopy forest in North-western Borneo in the Late Pleistocene: bones of Strategy I bats from the archaeological record of the Great Cave of Niah, Sarawak. *Palaeogeogr Palaeoclimatol Palaeoecol.* 2012;331-332:136-49.
62. Tosi AJ, Coke CS. Comparative phylogenetics offer new insights into the biogeographic history of *Macaca fascicularis* and the origin of the Mauritian macaques. *Mol Phylogenet Evol.* 2007;42:498-504.
63. Gathorne-Hardy FJ, Syaukani, Davies RG, Eggleton P, Jones DT. Quaternary rainforest refugia in southeast Asia: using termites (Isoptera) as indicators. *Biol J Linn Soc.* 2002;75:453-66.
64. O'Brien SJ, Mayr E. Bureaucratic mischief: recognizing endangered species and Science. 1991;251:1187-8.
65. Moritz C. Defining 'Evolutionary Significant Units' for conservation. *Tree.* 1994;9:373-5.
66. Kanthaswamy S, Kurushima JD, Smith DG. Inferring *Pongo* conservation units: A perspective based on microsatellite and mitochondrial DNA analyses. *Primates.* 2006;47:310-
67. Kanthaswamy S, Satkoski J, George D, Kou A, Erickson BJA, Smith DG. Hybridization and stratification of nuclear genetic variation in *Macaca mulatta* and *fascicularis*. *Int J Primatol.* 2008;29:1295-311.
68. Stevison LS, Kohn MH. Determining genetic background in captive stocks of cynomolgus macaques (*Macaca fascicularis*). *J Med Primatol* 2008;37:311-7.
69. Affara N, Bishop C, Brown W, Cooke H, Davey P, Ellis N, Graves JM, Jones M, Mitchell M, Rappold G, Tyler-Smith C, Yen P, Lau YFC. Report of the second international workshop on Y chromosome mapping 1995. *Cytogenet Cell Genet.* 1996;73:33-

70. Kim HS, Hirai H, Takenaka O. Molecular features of the TSPY gene of gibbons and Old World monkeys. *Chromosome Res.* 1996;4:500-6.
71. Glaser B, Grutzner F, Willman U, Stanyon R, Arnold N, Taylor K, Rietschel W, Zeitler S, Toder R, Schempp W. Simian Y chromosomes: Species-specific rearrangements of DAZ, RBM, and TSPY versus contiguity of PAR and SRY. *Mamm Genome.* 1998;9:226-
72. Drummond AJ, Ashton B, Buxton S, Cheung M, Cooper A, Duran C, Field M, Heled J, Kearse M, Markowitz S, Moir R, Stones-Havas S, Sturrock S, Thierer T, Wilson A. Geneious (Version 5.6.). 2012. <http://www.geneious.com>.
73. Swofford DL. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. 2003. Sinauer Associates, Sunderland, Massachusetts.
74. Tamura K, Stecher G, Peterson D, Filipinski A, Kumar S. MEGA: Molecular Evolutionary Genetics Analysis (Version 6.0). *Mol Biol Evol.* 2013;30:2725-9.
75. Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 2009;25:1451-2.
76. Nei M. *Molecular evolutionary genetics.* New York: Columbia University Press; 1987.
77. Kimura M. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol.* 1980;16:111-
78. Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution.* 1985;39:783-91.
79. Bouckaert R, Heled J, Kühnert D, Vaughan TG, Wu CH, Xie D, Suchard MA, Rambaut A, Drummond AJ. BEAST2: a software platform for Bayesian evolutionary analysis. *PLoS Comput Biol.* 2014;10:e100
80. Heled J, Drummond AJ. Calibrated tree priors for relaxed phylogenetics and divergence time estimation. *Syst Biol.* 2012;61:138-49.
81. Hasegawa M, Kishino H, Yano T. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J Mol Evol.* 1985;22:160-74.
82. Rambaut A, Suchard MA, Xie D, Drummond AJ. Tracer (Version 1.6) 2014. <http://beast.bio.ed.ac.uk/Tracer>.
83. Bandelt HJ, Forster P, Röhl A. Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol.* 1999;16:37-48.
84. Delson E. The oldest monkeys in Asia. In: Takenaka O, editor. *Evolution of Asian primates.* Inuyama, Aichi, Japan: International symposium conducted at Freude & Kyoto University, Primate Research Institute; 1996. p. 40.

Figures

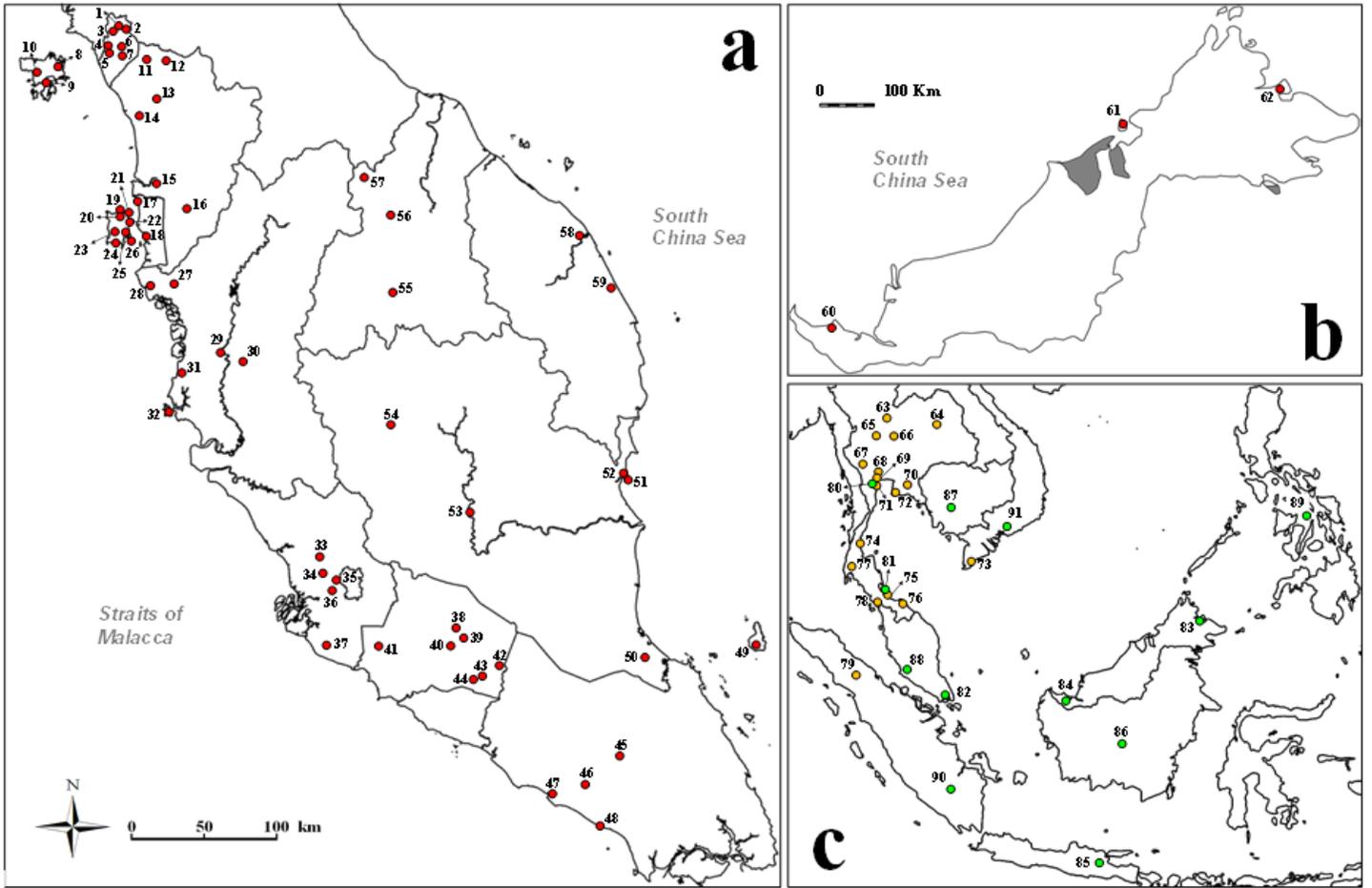


Figure 1

Map showing the localities of the *M. fascicularis* samples and sequences used in this study; (a) from Peninsular Malaysia, (b) from Malaysian Borneo (Sarawak and Sabah), and (c) downloaded sequences from previous studies as obtained from the GenBank (yellow circles represents sequences from Bunlungsup et al. [23] while green circles are sequences from Tosi, Morales and Melnick [9,15]). The location numbers in the map corresponds to that in Additional file 1: Table S1 (for samples obtained from this study) and Additional file 2: Table S2 (for sequences downloaded from GenBank)

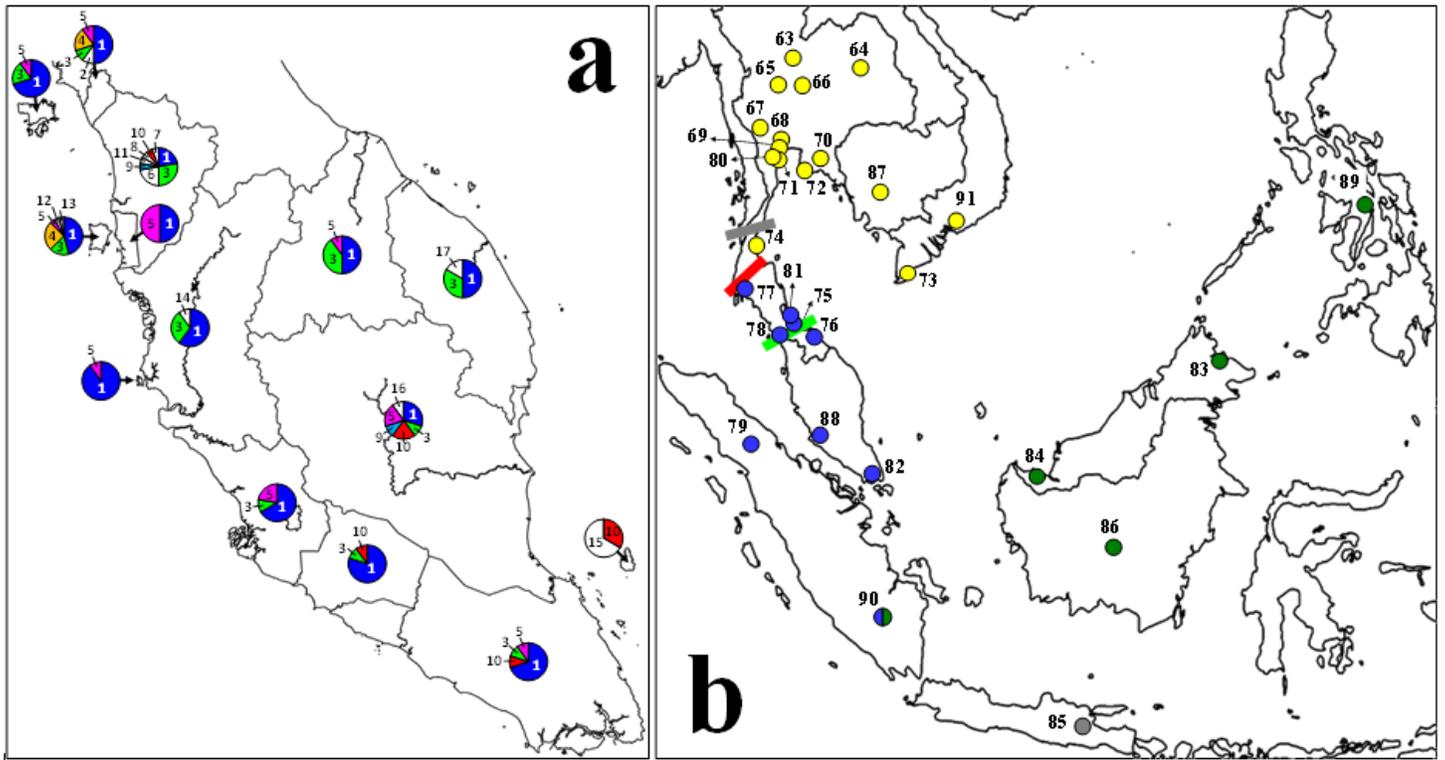


Figure 2

Y-chromosomal haplotypes mapping of *M. fascicularis*. (a) Haplotypes mapping (Haplotype 1 – 17) used in this study according to the states and major islands of Peninsular Malaysia. Sarawak and Sabah populations formed a single haplotype (Haplotype 18). The numbers correspond to the haplotype designation as listed in Additional file 1: Table S1. Note that only shared haplotypes were coloured. (b) Haplotypes mapping of sequences from the GenBank across Southeast Asia. Blue circles represent Haplotype 1, dark green circles represent Haplotype 18, grey circles represent Haplotype 19, and yellow circles represent Haplotype 20. Grey, red, and green lines indicate the approximate location of Isthmus of Kra, Surat Thani-Krabi depression, and Kangar-Pattani line, respectively. The location numbers in the map corresponds to that in Additional file 2: Table S2

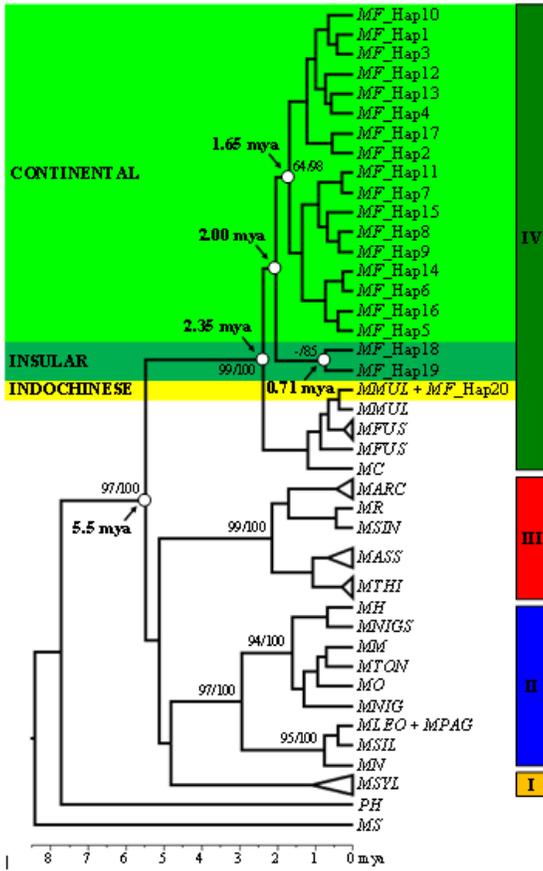


Figure 3

Y-chromosome phylogenetic tree constructed using the haplotypes of *Macaca* as represented by the BI analysis. ML analysis ($-\ln L = 3809.0846$) produced identical topology. Numbers on the branches represents support values for the ML and BI, respectively (only support values above 50% are shown). Roman numerals represent *Macaca* species groupings (I = sylvanus; II = silenus; III = sinica; IV = fascicularis). The species nomenclatures are abbreviated as shown in Additional file 2: Table S2

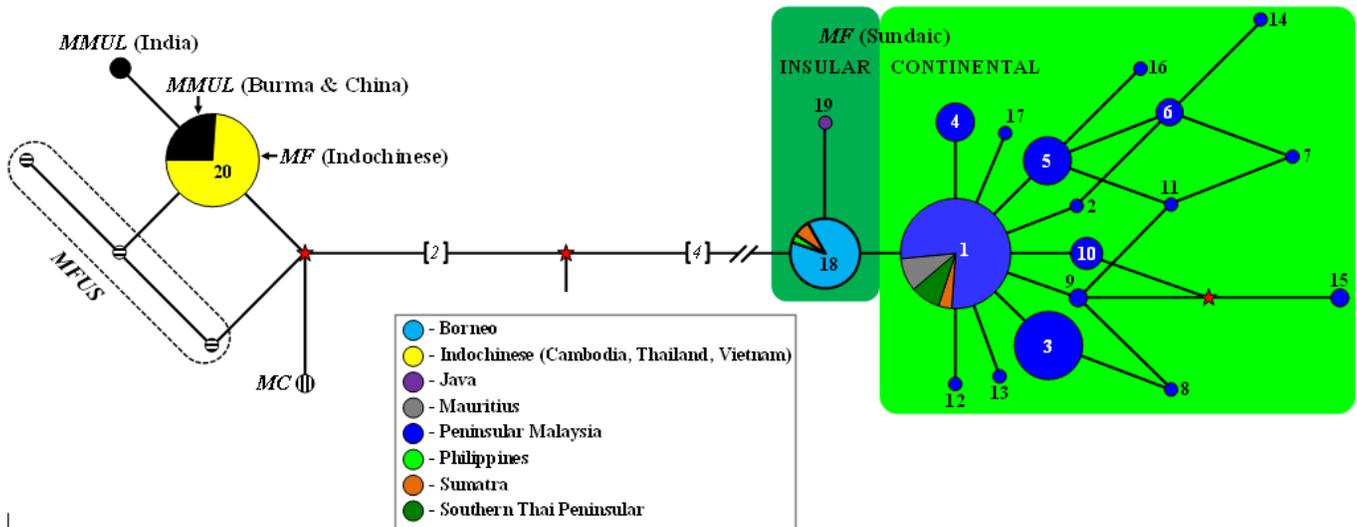


Figure 4

Haplotype median-joining network (MJN) constructed showing the relationships among the species within the fascicularis species group. The size of each circle is proportional to the number of individuals in each haplotype. The numbers on the nodes correspond to the haplotype designation as listed in Additional file 1 and 2: Table S1 and S2. The solid lines connecting the haplotype represent single mutations unless indicated otherwise (*italic numbers in parentheses*). Hypothetical haplotypes (not sampled) are represented by red stars

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Additonalfile1.xlsx](#)
- [Additonalfile2.xlsx](#)