

1
2
3
4 **Title: An expanded mammal mitogenome dataset from Southeast Asia**

5
6 Faezah Mohd Salleh^{1,2*}, Jazmín Ramos-Madrigal¹, Fernando Peñaloza^{3,4}, Shanlin Liu^{1,5}, Mikkol-Holger S.
7
8 Sinding^{1,6}, Riddhi P. Patel^{3,7}, Renata Martins³, Dorina Lenz³ Jörns Fickel^{3,8}, Christian Roos⁹, Mohd Shahir
9
10 Shamsir², Mohammad Shahfiz Azman¹⁰, Burton K. Lim¹¹, Stephen J. Rossiter¹², Andreas Wilting³, M.
11
12 Thomas P. Gilbert^{1,13*}

13
14 *Correspondence: faezah@fbb.utm.my and tgilbert@snm.ku.dk

15
16
17
18 ¹Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Øster
19
20 Voldgade 5-7, 1350, Copenhagen, Denmark

21
22 ²Faculty of Biosciences and Medical Engineering, Universiti Teknologi Malaysia, 81310 Johor Bahru,
23
24 Johor, Malaysia

25
26 ³Leibniz Institute for Zoo and Wildlife Research, Alfred-Kowalke Strasse 17, 10315 Berlin, Germany.

27
28 ⁴Undergraduate Program on Genomic Sciences, Universidad Nacional Autonoma de Mexico, 62210
29
30 Cuernavaca, Mexico

31
32 ⁵BGI-Shenzhen, Shenzhen, GuangDong, China

33
34 ⁶Natural History Museum, University of Oslo, P.O. Box 1172 Blindern, NO-0318, Oslo, Norway

35
36 ⁷Freie Universität Berlin, Kaiserswerther Str. 16-18, 14195 Berlin, Germany

37
38 ⁸University of Potsdam, Institute for Biochemistry and Biology, Karl-Liebknecht-Str 24-25, 14476
39
40 Potsdam, Germany

41
42 ⁹Gene Bank of Primates and Primate Genetics Laboratory, German Primate Center, Leibniz Institute for
43
44 Primate Research, Kellnerweg 4, 37077, Göttingen, Germany

45
46 ¹⁰Forest Biodiversity Division, Forest Research Institute Malaysia, 52109 Kepong, Selangor, Malaysia.

47
48 ¹¹Department of Natural History, Royal Ontario Museum, Toronto, Canada

49
50 ¹²School of Biological and Chemical Sciences, Queen Mary University of London, Mile End Road, London
51
52 E1 4NS, United Kingdom

53
54 ¹³NTNU University Museum, Norwegian University of Science and Technology, Trondheim, Norway

1
2
3
4
5
6 **Abstract**

7
8 **Background:**

9
10 Southeast (SE) Asia is one of the most biodiverse regions in the world and it holds approximately 20% of
11 all mammal species. Despite this, the majority of SE Asia's genetic diversity is still poorly characterised.

12
13
14 The growing interest in using environmental DNA (eDNA) to assess and monitor SE Asian species, in
15 particular threatened mammals - has created the urgent need to expand the available reference database
16 of mitochondrial barcode and complete mitogenome sequences. We have partially addressed this need
17 by generating 72 new mitogenome sequences reconstructed from DNA isolated from a range of historical
18 and modern tissue samples.
19

20
21
22 **Findings:**

23
24
25
26 Approximately 55 gigabases of raw sequence were generated. From this data we assembled 72 complete
27 mitogenome sequences, with an average depth of coverage of 102.9x and 55.2x for modern samples and
28 historical samples, respectively. This dataset represents 52 species, of which 30 species had no previous
29 mitogenome data available. The mitogenomes were geotagged to their sampling location, where known,
30 to display a detailed geographical distribution of the species.
31

32
33
34 **Conclusion:**

35
36
37
38 Our new database of 52 taxa will strongly enhance the utility of eDNA approaches for monitoring
39 mammals in SE Asia, as it greatly increases the likelihoods that identification of metabarcoding
40 sequencing reads can be assigned to reference sequences. This magnifies the confidence in species
41 detections and thus allows more robust surveys and monitoring programmes of SE Asia's threatened
42 mammal biodiversity. The extensive collections of historical samples from SE Asia in western and SE
43 Asian museums should serve as additional valuable material to further enrich this reference database.
44
45
46
47
48
49

50
51 **Keywords:**

52 invertebrate-derived (iDNA), metabarcoding, GenBank, Taxonomic assignment
53
54
55
56
57
58
59
60
61
62
63
64
65

1
2
3
4
5
6 **Data description:**
7

8 **Context**
9

10 Southeast (SE) Asia is one of the most biodiverse regions in the world, hosting ~20% of mammal species,
11 but is experiencing rapid deforestation for agriculture and development. To assess the ecological
12 consequences of land-use change, there is growing interest in using environmental DNA (eDNA) to
13 monitor mammal populations, particularly threatened taxa that often underpin conservation policies [1–4].
14 Yet current efforts are hampered by the lack of a reference database of mitochondrial barcodes and
15 complete mitogenome sequences. Currently there are 922 mammalian mitogenomes available in
16 Genbank. Unfortunately most are not tagged by location/origin. Data mining through manual screening of
17 each mitogenomes resulted in 174 terrestrial mammal species which are typical to SE Asia. In this work,
18 30 novel species are added, contributing to ~17% expansion of the current SE Asia mammal mitogenome
19 database.
20
21
22
23
24
25
26
27
28
29
30
31

32 **DNA extraction**
33

34 Genomic DNA was extracted from different sample types of 72 small mammals, comprising 52 species,
35 listed in Table 1 and Table 2. DNA from modern tissue and blood samples was isolated using the Qiagen
36 DNeasy extraction kit (Qiagen, Hilden, Germany, [QIAGEN, RRID:SCR_008539]) or Invitex DNA
37 extraction kit (Invitex GmbH, Berlin, Germany), as per standard protocols following the manufacturer's
38 guidelines. Historical samples obtained from the Zoological Museum, Natural History Museum of
39 Denmark, University of Copenhagen (ZM, KU) were treated differently according to type of tissue
40 (Additional file 1a), while at the German Primate Center DNA extraction from museum specimens
41 followed Liedigk et al. (2015) [5] using the Gen-IAL First All tissue kit (Gen-IAL, Troisdorf, Germany).
42 Complete details of sample information are provided in Additional file 2.
43
44
45
46
47
48
49
50
51
52
53
54

55 **Data Validation and Quality Control**
56

57 **Mitogenome sequencing, assembly and annotation**
58

59 Mitogenomes were generated using several approaches. In Copenhagen, author FMS constructed
60
61
62
63
64
65

1
2
3
4 Illumina shotgun libraries with insert sizes ranging between 50 bp to 400 bp. To construct libraries, DNA
5
6 was sheared to the target size range using Bioruptor® XL (Diagenode, USA [Diagenode,
7
8 RRID:SCR_014807]) and converted into an Illumina-compatible sequencing library using the NEBNext
9
10 E6070 kit (New England Biolabs, UK). The libraries were PCR amplified with index primers and purified
11
12 using Qiaquick columns (Qiagen, Hilden, Germany) according to the manufacturer's instruction
13
14 (Additional file 1b). Multiple libraries were combined together into three pools, normalized to 10 nM and
15
16 sequenced across three lanes of Illumina HiSeq 2500 using SR100bp chemistry. In Berlin and
17
18 Goettingen, mitogenomes were generated by author PRP and CR using overlapping PCR products using
19
20 long range PCR (Additional file 1c) followed by library construction and MiSeq sequencing, or Sanger
21
22 sequencing as described in [6] and [5, 7, 8], respectively. Author RM's mitogenomes were done using
23
24 methods outlined in [9]. Further details about laboratory methods are described in Additional file 1.
25
26
27
28
29

30
31 Raw reads for FMS samples were assembled independently by authors FMS and FP using two different
32
33 approaches, then compared for consistency. Author FMS trimmed the reads for sequencing adapters, low
34
35 quality stretches and leading/trailing Ns using AdapterRemoval 1.2 (AdapterRemoval,
36
37 RRID:SCR_011834) [10]. The mitochondrial genome was reconstructed with MITObim v1.8 [11] using the
38
39 reference mitogenome of the closest species available in GenBank as the seed reference (Additional file
40
41 2). In order to obtain the mapping statistics of the samples, we ran PALEOMIX v 1.2.6 [12] with default
42
43 parameters where reads shorter than 25 bp after trimming were discarded. The trimmed reads were
44
45 aligned against the newly assembled mitogenome generated by MITObim using Burrows–Wheeler
46
47 Aligner (BWA) [13]. Alignments showing low-quality scores and PCR duplicates were further removed
48
49 using the MarkDuplicates program from Picard tools, and reads were locally realigned around small
50
51 insertions and deletions (indels) to improve overall genome quality using the IndelRealigner tool from the
52
53 Genome Analysis Toolkit (GATK), (GATK, RRID:SCR_001876) [14]. In contrast, author FP inputted
54
55 the trimmed reads into mitoMaker [15] which performs a de-novo and reference-based assembly using
56
57 SOAPdenovoTrans v1.03 (SOAPdenovo-Trans, RRID:SCR_013268) [16] and MITObim v1.7 [11]. Post
58
59
60
61
62
63
64
65

1
2
3
4 assembly, the FMS and FP mitogenomes were manually compared for consistency by FMS to generate
5
6 the final consensus sequences. These assemblies were automatically annotated using tRNAscan-SE
7
8 v1.4 (tRNAscan-SE, RRID:SCR_010835) [17] and BLAST v2.2.29 (NCBI BLAST, RRID:SCR_004870)
9
10 [18] using the mitochondrial genomes found in the NCBI RefSeq (RefSeq, RRID:SCR_003496) [19] as
11
12 references.
13

14
15
16 For mitogenome constructed by author RM, Illumina sequence reads were de-multiplexed according to
17
18 the respective indexes with the Illumina software bcl2fastq v2.17 (Illumina, San Diego, CA, USA) and
19
20 adapters were clipped from the sequence reads with the software cutadapt v1.3 [20]. Quality trimming
21
22 was done through a sliding window approach (10 bp; Q20) and all reads shorter than 20 bp were
23
24 removed from the analyses. Mitogenome references from target or closely related species were used for
25
26 mapping of the sequencing reads. Aligned reads were de-duplicated using MarkDuplicates from Picard-
27
28 tools v1.106 (Picard, RRID:SCR_006525) (<https://github.com/broadinstitute/picard>). VariantCalling was
29
30 carried out using Samtools v1.1 (SAMTOOLS, RRID:SCR_002105)[13] and Bcftools v1.2
31
32 (SAMtools/BCFtools, RRID:SCR_005227) [21] For each sample, GATK [14] Variant Calling output files
33
34
35 were further filtered to have a minimum read coverage $\geq 3x$, and variants were only called when the
36
37
38 corresponding base was represented by $\geq 50\%$, otherwise this position was "N"-masked.
39
40
41

42
43
44 Numbers of raw reads generated for each samples and mapping statistics for all 72 mitogenome
45
46 assemblies are shown in Additional file 2. Sanger sequenced mitogenomes were checked with 4Peaks
47
48 1.8 (4Peaks, RRID:SCR_000015) [22], assembled with SeaView 4.5.4 [23] and annotated with DOGMA
49
50 [24]. All mitogenomes were checked manually by eye to identify possible errors caused by insertion and
51
52 deletions in Tablet [25]. The final mitochondrial genomes have been uploaded to GenBank (accession
53
54 numbers are provided in Table 1 and Table 2). The details of all new mitogenomes assembled in this
55
56 work are given in Table 1 and Table 2. Mitogenomes (60 samples) with known localities were geotagged
57
58 and mapped to display its geographical distribution (Figure 1).
59
60
61
62
63
64
65

1
2
3
4
5
6 **Phylogenetic analysis**
7

8 All the sequenced mitogenomes were aligned using MAFFT v7.158b (MAFFT, RRID:SCR_011811) [26],
9
10 using the E-INS-i option (Additional file 3). RAxML v8.0.26 (RAxML, RRID:SCR_006086) [27] was used to
11
12 perform the phylogenetic analysis with a GTR+GAMMA model of nucleotide substitution. To obtain node
13
14 support, we used 100 bootstrap pseudo-replicates (Figure 2). The newick file is provided as Additional file
15
16 4.
17
18
19

20 **Re-use Potential**
21

22 We anticipate that the now expanded mitogenome reference dataset for SE Asian mammals will provide
23
24 benefits for a number of research areas. Firstly it should enhance the power of eDNA and other
25
26 metabarcoding/barcoding approaches that relate to the identification of SE asian mammals, by conferring
27
28 the ability to identify more species to the species level. This in turn has practical applications for those
29
30 monitoring SE Asia's threatened mammal biodiversity, combatting trade in mammal species and so on.
31
32 Secondly the data will also have relevance to phylogenetic and population studies based on mtDNA data,
33
34 which will be of use as we investigate the evolutionary history of this biodiversity hotspot.
35
36
37
38

39 **Availability of supporting data**
40

41 Raw shotgun data are deposited in the SRA under bioproject number PRJNA361218 and is available in the
42
43 GigaScience repository, GigaDB [28].
44
45

46 **Abbreviations**
47

48 SE: South East; eDNA: environmental deoxyribonucleic acid; bp: basepair; PCR: polymerase chain
49
50 reaction; BWA: Burrows–Wheeler Aligner; GATK: Genome Analysis Toolkit; BLAST: Basic Local
51
52 Alignment Search Tool; NCBI RefSeq: National Center for Biotechnology Information Reference
53
54 Sequence Database; MAFFT: Multiple Alignment using Fast Fourier Transform ; RAxML: Randomized
55
56 Axelerated Maximum Likelihood.
57
58
59
60
61
62
63
64
65

1
2
3
4 **Competing interests**

5
6 The authors declare that they have no competing interests.
7
8
9

10 **Funding**

11
12 This project was funded by the Malaysian Government (FMS), Lundbeck Foundation grant R52-5062
13 (MTPG), Leibniz-Association grant SAW-2013-IZW-2 (JF), the German Federal Ministry of Education and
14 Research grant BMBF FKZ: 01LN1301A (AW) and the German Primate Center (CR).
15
16
17
18
19

20 **Authors' contributions**

21
22 FMS, AW, JF and MTPG conceived the project. FMS, MHSS, MSS, MSA, RM, PRP, CR, BKL, and SJR
23 collected the samples and extracted the genomic DNA. FMS, RM, PRP and CR constructed the libraries
24 and did sequencing. FMS, JRM, FP, SL, PRP, RM, DL and CR assembled the mitogenomes and
25 performed mitogenome analysis. FMS, SL, PRP and MTPG wrote the article. All authors discussed the
26 project and data. All authors read and approved the final manuscript.
27
28
29
30
31
32
33
34

35 **Acknowledgements**

36
37 We thank the Danish National High-throughput Sequencing Centre for assistance in generating Illumina
38 data. We also thank the staff of the zoos in Hannover, Copenhagen, Gettorf, Belfast, Bristol, Howletts
39 Wild Animal Park, the staff of the Ludwig-Maximilians-University Munich, the Bavarian State Collection
40 Munich, the National Museums Scotland, Edinburgh, the Natural History Museum Berlin, Naturalis
41 Leiden, and the Zoological Museum, Natural History Museum of Denmark, University of Copenhagen for
42 providing samples for this work. BKL gratefully acknowledges the Indonesian and Malaysian government
43 authorities for issuing scientific permits.
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58

59 **Author details**

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

¹Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Øster Voldgade 5-7, 1350, Copenhagen, Denmark. ²Faculty of Biosciences and Medical Engineering, Universiti Teknologi Malaysia, 81310 Johor Bahru, Johor, Malaysia. ³Leibniz Institute for Zoo and Wildlife Research, Alfred-Kowalke Strasse 17, 10315 Berlin, Germany. ⁴Undergraduate Program on Genomic Sciences, Universidad Nacional Autonoma de Mexico, 62210 Cuernavaca, Mexico. ⁵BGI-Shenzhen, Shenzhen, Guangdong, China. ⁶Natural History Museum, University of Oslo, P.O. Box 1172 Blindern, NO-0318, Oslo, Norway. ⁷Freie Universität Berlin, Kaiserswerther Str. 16-18, 14195 Berlin, Germany. ⁸University of Potsdam, Institute for Biochemistry and Biology, Karl-Liebknecht-Str 24-25, 14476 Potsdam, Germany. ⁹Gene Bank of Primates and Primate Genetics Laboratory, German Primate Center, Leibniz Institute for Primate Research, Kellnerweg 4, 37077, Göttingen, Germany. ¹⁰Forest Biodiversity Division, Forest Research Institute Malaysia, 52109 Kepong, Selangor, Malaysia. ¹¹Department of Natural History, Royal Ontario Museum, Toronto, Canada. ¹²School of Biological and Chemical Sciences, Queen Mary University of London, Mile End Road, London E1 4NS, United Kingdom. ¹³NTNU University Museum, Norwegian University of Science and Technology, Trondheim, Norway.

References

1. Bohmann K, Evans A, Gilbert MTP, Carvalho GR, Creer S, Knapp M, Yu DW, de Bruyn M (2014) Environmental DNA for wildlife biology and biodiversity monitoring. *Trends Ecol Evol* 29:358–367
2. Lee P-S, Gan HM, Clements GR, Wilson J-J (2016) Field calibration of blowfly-derived DNA against traditional methods for assessing mammal diversity in tropical forests. *Genome* 59:1008–1022
3. Schnell IB, Sollmann R, Calvignac-Spencer S, Siddall ME, Douglas WY, Wilting A, Gilbert MTP (2015) iDNA from terrestrial haematophagous leeches as a wildlife surveying and monitoring tool--prospects, pitfalls and avenues to be developed. *Front Zool* 12:24
4. Schnell IB, Thomsen PF, Wilkinson N, Rasmussen M, Jensen LRD, Willerslev E, Bertelsen MF, Gilbert MTP (2012) Screening mammal biodiversity using DNA from leeches. *Curr Biol* 22:R262–3
5. Liedigk R, Kolleck J, Böker KO, et al (2015) Mitogenomic phylogeny of the common long-tailed macaque (*Macaca fascicularis fascicularis*). *BMC Genomics* 16:222
6. Patel RP, Förster DW, Kitchener AC, et al (2016) Two species of Southeast Asian cats in the genus *Catopuma* with diverging histories: an island endemic forest specialist and a widespread habitat generalist. *Open Science* 3:160350
7. Roos C, Zinner D, Kubatko LS, et al (2011) Nuclear versus mitochondrial DNA: evidence for hybridization in colobine monkeys. *BMC Evol Biol* 11:77
8. Liedigk R, Roos C, Brameier M, Zinner D (2014) Mitogenomics of the Old World monkey tribe Papionini. *BMC Evol Biol* 14:176
9. Fortes GG, Paijmans JLA (2015) Analysis of Whole Mitogenomes from Ancient Samples. *Methods Mol Biol* 1347:179–195
10. Lindgreen S (2012) AdapterRemoval: easy cleaning of next-generation sequencing reads. *BMC Res Notes* 5:337
11. Hahn C, Bachmann L, Chevreur B (2013) Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. *Nucleic Acids Res.*
12. Schubert M, Ermini L, Der Sarkissian C, et al (2014) Characterization of ancient and modern genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX. *Nat Protoc* 9:1056–1082
13. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25:1754–1760
14. McKenna A, Hanna M, Banks E, et al (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20:1297–1303
15. mitoMaker. <https://sourceforge.net/projects/mitomaker/>. Accessed 26 May 2017
16. Xie Y, Wu G, Tang J, et al (2014) SOAPdenovo-Trans: de novo transcriptome assembly with short RNA-Seq reads. *Bioinformatics* 30:1660–1666
17. Lowe TM, Eddy SR (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 25:955–964

- 1
- 2
- 3
- 4 18. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL (2009) BLAST+:
5 architecture and applications. *BMC Bioinformatics* 10:421
- 6
- 7 19. RefSeq: NCBI Reference Sequence Database. <https://www.ncbi.nlm.nih.gov/refseq>. Accessed 26
8 May 2017
- 9
- 10 20. Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads.
11 *EMBnet.journal* 17:10–12
- 12
- 13 21. bcftools. <https://github.com/samtools/bcftools>. Accessed 26 May 2017
- 14
- 15 22. Nucleobytes: software for science. <http://nucleobytes.com>. Accessed 26 May 2017
- 16
- 17 23. Gouy M, Guindon S, Gascuel O (2010) SeaView version 4: A multiplatform graphical user interface
18 for sequence alignment and phylogenetic tree building. *Mol Biol Evol* 27:221–224
- 19
- 20 24. Wyman SK, Jansen RK, Boore JL (2004) Automatic annotation of organellar genomes with DOGMA.
21 *Bioinformatics* 20:3252–3255
- 22
- 23 25. Milne I, Stephen G, Bayer M, Cock PJA, Pritchard L, Cardle L, Shaw PD, Marshall D (2012) Using
24 Tablet for visual exploration of second-generation sequencing data. *Brief Bioinform* bbs012
- 25
- 26 26. Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements
27 in performance and usability. *Mol Biol Evol* 30:772–780
- 28
- 29 27. Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large
30 phylogenies. *Bioinformatics* 30:1312–1313
- 31
- 32 28. Salleh FM, Ramos-Madriral J, Penaloza F, Liu S, Sinding MS, Patel RP, Martins R, Lenz D, Fickel J,
33 Roos C, Shamsir MS, Azman MS, Lim BK, Rossiter SJ, Wilting A, Gilbert MP. Supporting data for "An
34 expanded mammal mitogenome dataset from Southeast Asia" GigaScience Database. 2017.
35 <http://dx.doi.org/10.5524/100313>
- 36
- 37
- 38
- 39
- 40
- 41
- 42
- 43
- 44
- 45

46 **Tables and Figures**

- 47
- 48 1. Table 1: List of mitogenomes assembled in this work that supplement pre-existing mitogenome
49 references already available in GenBank
- 50
- 51
- 52 2. Table 2: List of mitogenomes assembled in this work that have no previous complete mitogenome
53 reference available in GenBank
- 54
- 55
- 56 3. Figure 1: Geographical distribution of mitogenomes assembled in this work (60 mitogenomes with
57 known locality)
- 58
- 59
- 60
- 61
- 62
- 63
- 64
- 65

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

4. Figure 2: Phylogenetic tree of mitogenomes assembled in this work

Additional files

1. Additional file 1: DNA extraction of historical samples, library construction and primer information
2. Additional file 2: Sample information sheet of mitogenomes assembled in this work
3. Additional file 3: Alignment of mitogenomes assembled in this work
4. Additional file 4: Newick file for phylogenetic tree

Table 1: List of mitogenomes assembled in this work that supplement pre-existing mitogenome references already available in GenBank

No	GenBank ID	Common name	Genus	Species	Assembly Size	Locality	Source	Sample date of collection	Data by
1	KY117537	Hog deer	<i>Axis</i>	<i>porcinus</i>	16402	CPH Zoo	ZM, KU	21/8/1912	FMS/FP
2	KY117538	Pallas's squirrel	<i>Callosciurus</i>	<i>erythraeus</i>	16656	Bangkok, Thailand	ZM, KU	25/5/1969	FMS/FP
3	KX265095	Bay Cat	<i>Catopuma</i>	<i>badia</i>	16960	Sabah, Malaysia	National Museum Scotland	20/04/2000	PRP
4	KX224524	Asiatic Golden Cat	<i>Catopuma</i>	<i>temminckii</i>	16,960	Thailand	American Museum of National History, New York.	10/10/1927	PRP
5	KY117545	Sumatran Rhino	<i>Dicerorhinus</i>	<i>sumatrensis</i>	16466	Sumatra, Indonesia	Naturalis, Leiden, The Netherlands	1880	RM
6	KY117546	Least pygmy squirrel	<i>Exilisciurus</i>	<i>exilis</i>	16637	Indonesia	ROM	16/06/1993	FMS/FP
7	KY117548	Hose's mongoose	<i>Herpestes</i>	<i>javanicus</i>	16340	Java, Indonesia	ZM, KU	12/3/1947	FMS/FP
8	KY117550	Three-striped ground Squirrel	<i>Lariscus</i>	<i>indsignis</i>	16399	Maybe Malaysia	ZM, KU	unknown	FMS/FP
9	KY117592	Black Crested Macaque	<i>Macaca</i>	<i>nigra</i>	16558	captive	Gettorf Zoo, Germany	18/07/2000	CR
10	KY117593	Northern Pig-tailed Macaque	<i>Macaca</i>	<i>leonina</i>	16554	captive	Ludwig-Maximilians-University Munich, Germany	6/3/1995	CR
11	KY117594	Southern pig-tailed macaque	<i>Macaca</i>	<i>nemestrina</i>	16531	Peninsular Malaysia	National Museum Scotland, Edinburgh, UK	unknown	CR
12	KT288227	Marbled Cat	<i>Pardofelis</i>	<i>marmorata</i>	17218	Sumatra, Indonesia	National Archaeological Museum of the Netherlands, Leiden	30/08/1930	PRP
13	KY117602	Sumatra Surili	<i>Presbytis</i>	<i>melalophos</i>	16558	captive	Howletts Wild Animal Park, UK	23/7/1999	CR
14	KR135743	Flat-headed Cat	<i>Prionailurus</i>	<i>planiceps</i>	17704	Sabah, Malaysia	Sabah wildlife department	25/04/2000	PRP
15	KY117580	Malayan Field Rat	<i>Rattus</i>	<i>tiomanicus</i>	16415	SPF Bidor, Perak, Malaysia	FRIM	12/2/2011	FMS/FP
16	KY117579	Malayan Field Rat	<i>Rattus</i>	<i>tiomanicus</i>	16312	Indonesia	ROM	01/06/1993	FMS/FP
17	KY117581	Malayan Field Rat	<i>Rattus</i>	<i>tiomanicus</i>	16305	Hutan Simpan Chikus, Tapah Perak, Malaysia	FRIM	13/1/2011	FMS/FP
19	KY117582	Black Giant Squirrel	<i>Ratufa</i>	<i>bicolor peninsulae</i>	16600	Bang Nara, Malakka, Thailand	ZM, KU	3/12/1932	FMS/FP

18	KY117574	Javan Rhino	<i>Rhinoceros</i>	<i>sondaicus</i>	16417	Java, Indonesia	Copenhagen Natural History Museum	unknown	RM
20	KY117575	Javan Rusa	<i>Rusa</i>	<i>timorensis</i>	16437	Toeloeng Agoeng, West Java, Indonesia	Naturalis, Leiden, The Netherlands	unknown	RM
21	KY117576	Indian Sambar Deer	<i>Rusa</i>	<i>unicolor dejeani</i>	16437	Mentawai, Indonesia	Naturalis, Leiden, The Netherlands	unknown	RM
22	KY117599	Western Purple-faced Langur	<i>Semnopithecus</i>	<i>vetulus</i>	16545	captive	Belfast Zoo, UK	9/11/1998	CR
23	KY117589	Malayan Tapir	<i>Tapirus</i>	<i>indicus</i>	16794	captive	Copenhagen Zoo	11/1/2015	FMS/FP
24	KY117598	Silvered Langur	<i>Trachypithecus</i>	<i>cristatus</i>	16551	North Sumatra, Indonesia	Bavarian State Collection Munich, Germany	1911	CR

ZM, KU: Zoological Museum, University of Copenhagen; ROM: Royal Ontario Museum; FRIM: Forest Research Institute, Malaysia

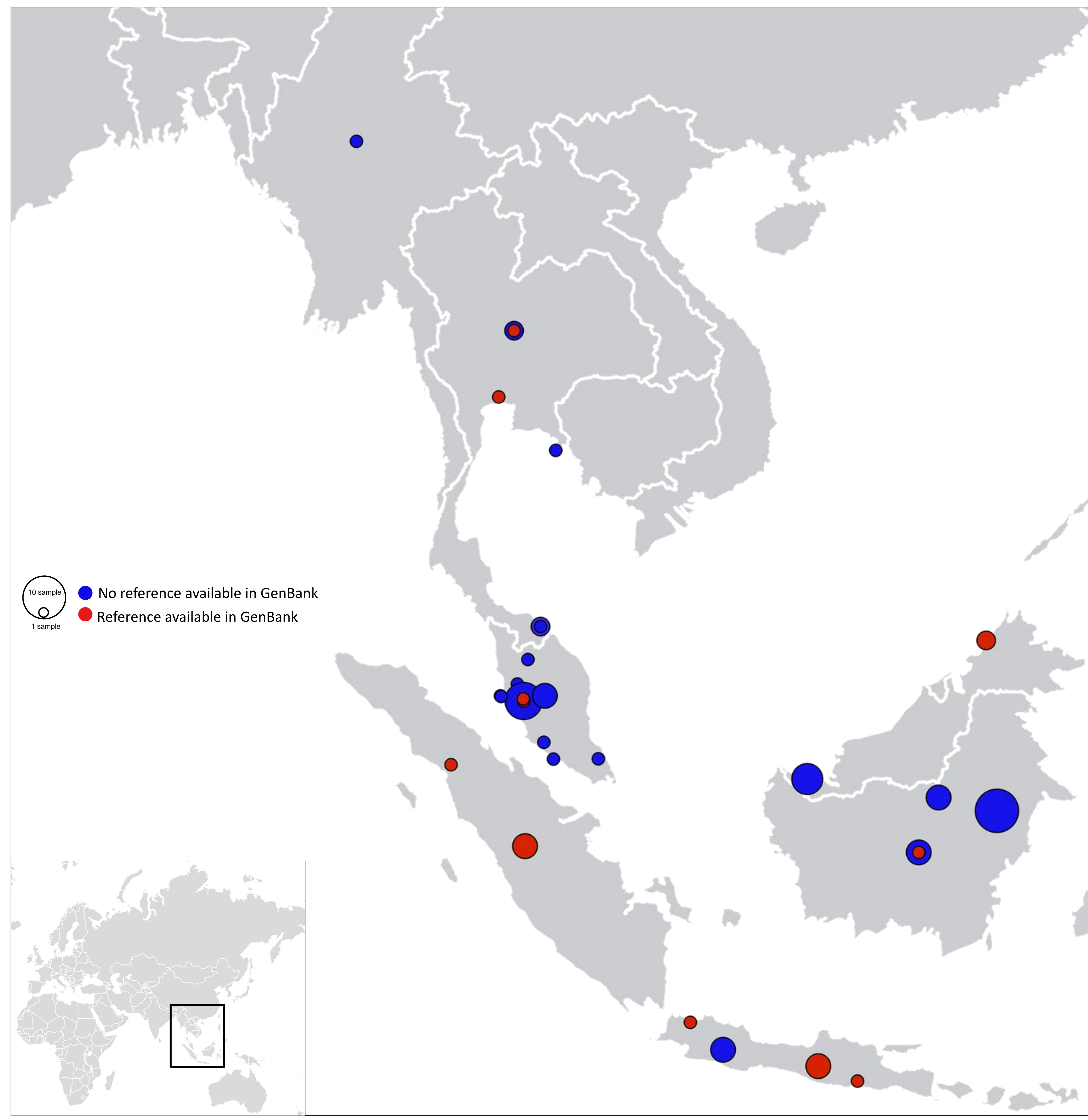
Table 2: List of mitogenomes assembled in this work that have no previous complete mitogenome reference available in GenBank

No	GenBank ID	Common name	Genus	Species	Assembly Size	Locality	Source	Sample date of collection	Data by
1	KY117536	Asian Small-clawed Otter	<i>Aonyx</i>	<i>cinereus</i>	16153	captive	Copenhagen Zoo	08/08/11	FMS/FP
2	KY117535	Asian Small-clawed Otter	<i>Aonyx</i>	<i>cinereus</i>	16153	Sarawak, Malaysia	British Museum of Natural History, London	25/8/2010	FMS/FP
3	KY117560	Binturong	<i>Arctictis</i>	<i>binturong</i>	17067	unknown	Tierpark, Berlin	29/11/2010	PRP
4	KY117541	Plantain Squirrel	<i>Callosciurus</i>	<i>notatus</i>	16582	Hutan Bidor, Perak, Malaysia	FRIM	11/2/2011	FMS/FP
5	KY117542	Plantain squirrel	<i>Callosciurus</i>	<i>notatus</i>	16599	East Kalimantan, Indonesia	ROM	03/06/1993	FMS/FP
6	KY117543	Prevost's squirrel	<i>Callosciurus</i>	<i>prevostii</i>	16674	East Kalimantan, Indonesia	ROM	15/06/1993	FMS/FP
7	KY117540	Variable squirrel	<i>Callosciurus</i>	<i>finlaysonii frandseni</i>	15747	Koh Chang, Thailand	ZM, KU	14/1/1900	FMS/FP
8	KY117539	Variable squirrel	<i>Callosciurus</i>	<i>finlaysonii</i>	16489	Central Thailand	ZM, KU	2/2/1928	FMS/FP
9	KY117544	Sunda Otter Civet	<i>Cynogale</i>	<i>bennetti</i>	15784	Borneo	British Museum of Natural History, London	25/8/2010	FMS/FP
10	KY117549	Greater Mouse deer	<i>Tragulus</i>	<i>napu</i>	15778	Bang Nara, Thailand	ZM, KU	11/10/1931	FMS/FP
11	KY117552	Long-tailed Giant Rat	<i>Leopoldamys</i>	<i>sabanus</i>	15973	G. Telapak Buruk, Negeri Sembilan, Malaysia	FRIM	24/2/2010	FMS/FP
12	KY117553	Long-tailed Giant Rat	<i>Leopoldamys</i>	<i>sabanus</i>	15972	Teluk Segadas, P. Pangkor, Perak, Malaysia	FRIM	19/3/2010	FMS/FP
13	KY117554	Long-tailed Giant Rat	<i>Leopoldamys</i>	<i>sabanus</i>	15974	Hutan Simpan Temengor, Gerik Perak, Malaysia	FRIM	23/1/2014	FMS/FP
14	KY117555	Long-tailed Giant Rat	<i>Leopoldamys</i>	<i>sabanus</i>	15972	Hutan Simpan Lenggong, Kluang, Johor, Malaysia	FRIM	19/2/2014	FMS/FP
15	KY117551	Long-tailed giant rat	<i>Leopoldamys</i>	<i>sabanus</i>	15974	Malaysia	ROM	28/05/1993	FMS/FP
16	KY117556	Hairy-nosed otter	<i>Lutra</i>	<i>sumatrana</i>	16580	Bang Nara, Thailand	ZM, KU	1/4/1939	FMS/FP
17	KY117557	Smooth-coated otter	<i>Lutrogale</i>	<i>perspicillata</i>	16042	Melaka, Malaysia	British Museum of Natural History, London	25/8/2010	FMS/FP

18	KY117558	Smooth-coated otter	<i>Lutrogale</i>	<i>perspicillata</i>	16041	Bang Nara, Thailand	ZM, KU	24/1/1933	FMS/FP
19	KY117591	Moor macaque	<i>Macaca</i>	<i>maura</i>	16563	captive	Hannover Zoo, Germany	20/8/1998	CR
20	KY117564	Rajah/Brown Spiny Rat	<i>Maxomys</i>	<i>rajah</i>	16200	Indonesia	ROM	06/06/1993	FMS/FP
21	KY117562	Rajah/Brown Spiny Rat	<i>Maxomys</i>	<i>rajah</i>	16296	Teluk Segadas, P. Pangkor, Perak, Malaysia	FRIM	19/3/2010	FMS/FP
22	KY117563	Rajah/Brown Spiny Rat	<i>Maxomys</i>	<i>rajah</i>	16296	Pasir Bogak, P. Pangkor, Perak, Malaysia	FRIM	18/3/2010	FMS/FP
23	KY117567	Red Spiny Rat	<i>Maxomys</i>	<i>surifer</i>	16286	50 ha, Pasoh, Negeri Sembilan, Malaysia	FRIM	12/6/2008	FMS/FP
24	KY117566	Red Spiny Rat	<i>Maxomys</i>	<i>surifer</i>	16290	Indonesia	ROM	21/05/1993	FMS/FP
25	KY117565	Red Spiny Rat	<i>Maxomys</i>	<i>surifer</i>	16286	Malaysia	ROM	17/05/2013	FMS/FP
26	KY117570	Whitehead's Spiny Rat	<i>Maxomys</i>	<i>whiteheadi</i>	16316	Hutan Simpan Bikam, Perak, Malaysia	FRIM	12/2/2011	FMS/FP
27	KY117571	Whitehead's Spiny Rat	<i>Maxomys</i>	<i>whiteheadi</i>	16316	Keruing Trail, FRIM, Kepong, Selangor, Malaysia	FRIM	13/3/2013	FMS/FP
28	KY117568	Whitehead's Spiny Rat	<i>Maxomys</i>	<i>whiteheadi</i>	16287	Hutan Simpan Bikam, Perak, Malaysia	FRIM	12/2/2011	FMS/FP
29	KY117569	Whitehead's Spiny Rat	<i>Maxomys</i>	<i>whiteheadi</i>	16429	Bukit Tapah, Perak, Malaysia	FRIM	23/3/2011	FMS/FP
30	KY052142	Indian muntjac	<i>Muntiacus</i>	<i>muntjak</i>	16354	West Java, Indonesia	Vienna NHM	1858	RM
31	KY117559	Bornean yellow muntjac	<i>Muntiacus</i>	<i>atherodes</i>	16354	Koemai, West Borneo	Bonn NHM	1938	RM
32	KY117573	Dark-tailed Tree Rat	<i>Niviventer</i>	<i>cremoriventer</i>	16322	Track 5 (G.Inas), Kedah, Malaysia	FRIM	5/11/2009	FMS/FP
33	KY117572	Dark-tailed tree rat	<i>Niviventer</i>	<i>cremoriventer</i>	16234	Malaysia	ROM	17/05/2013	FMS/FP
34	KY117600	Grizzled Leaf Monkey	<i>Presbytis</i>	<i>comata comata</i>	16551	captive	Howletts Wild Animal Park, UK	23/12/1999	CR
35	KY117601	Mitred Leaf Monkey	<i>Presbytis</i>	<i>mitrata</i>	16555	captive	Howletts Wild Animal Park, UK	12/11/1998	CR
36	KX857784	Leopard cat	<i>Prionailurus</i>	<i>bengalensis</i>	16989	Thailand	American Museum of National History, New York.	25/02/1924	PRP

37	KY117578	Annandale's Sundaic Rat	<i>Rattus</i>	<i>annandalei</i>	16297	Hutan Simpan Bikam, Perak, Malaysia	FRIM	12/2/2011	FMS/FP
38	KY117577	Annandale's Sundaic Rat	<i>Rattus</i>	<i>annandalei</i>	16301	Hutan Simpan Bikam, Perak, Malaysia	FRIM	11/2/2011	FMS/FP
39	KY117583	Mountain giant Sunda rat	<i>Sundamys</i>	<i>infraluteus</i>	16297	Malaysia	ROM	18/05/2013	FMS/FP
40	KY117585	Müller's Giant Sunda Rat	<i>Sundamys</i>	<i>meulleri</i>	16326	Track 1 (G.Inas), Kedah, Malaysia	FRIM	5/11/2009	FMS/FP
41	KY117584	Müller's giant Sunda rat	<i>Sundamys</i>	<i>meulleri</i>	16304	Malaysia	ROM	01/06/2013	FMS/FP
42	KY117586	Brooke's squirrel	<i>Sundasciurus</i>	<i>brookei</i>	16417	East Kalimantan, Indonesia	ROM	13/06/1993	FMS/FP
43	KY117587	Low's squirrel	<i>Sundasciurus</i>	<i>lowii</i>	16307	East Kalimantan, Indonesia	ROM	06/06/1993	FMS/FP
44	KY117588	Low's squirrel	<i>Sundasciurus</i>	<i>sp</i>	16458	East Kalimantan, Indonesia	ROM	21/06/1993	FMS/FP
45	KY117595	Phayre's Langur	<i>Trachypithecus</i>	<i>phayrei phayrei</i>	16548	South West Myanmar	Natural History Museum Berlin, Germany	unknown	CR
46	KY117596	East Javan Ebony Langur	<i>Trachypithecus</i>	<i>auratus</i>	16552	captive	Bristol Zoo, UK	26/10/2010	CR
47	KY117597	West Javan Ebony Langur	<i>Trachypithecus</i>	<i>mauritus</i>	16554	West Java, Indonesia	Naturalis Leiden; Netherlands	unknown	CR
48	KY117590	Long-tailed porcupine	<i>Trichys</i>	<i>fasciculata</i>	16328	Borneo	ZM, KU	5/10/1894	FMS/FP

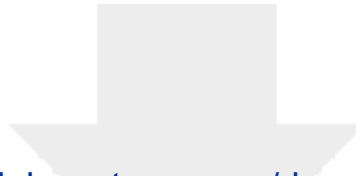
ZM, KU: Zoological Museum, University of Copenhagen; ROM: Royal Ontario Museum; FRIM: Forest Research Institute, Malaysia; NHM: Natural History Museum





Click here to access/download
Supplementary Material
READMEsallehetal.txt

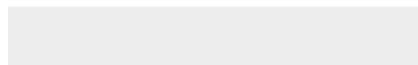


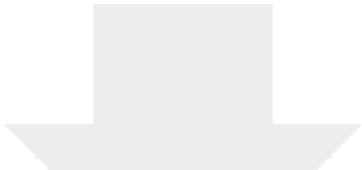


Click here to access/download

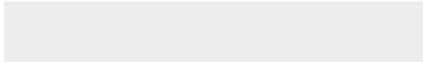

Supplementary Material

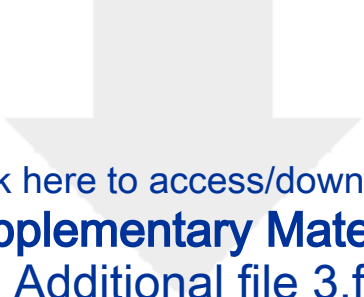
1DatainformationSallehetal.docx



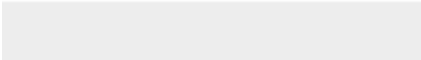


Click here to access/download
Supplementary Material
5 Additional file 2.xlsx



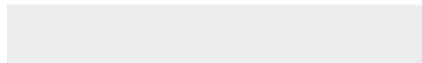


Click here to access/download
Supplementary Material
5 Additional file 3.fa





Click here to access/download
Supplementary Material
5 Additional file 4 newick.txt

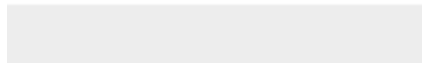




[Click here to access/download](#)

Supplementary Material

5Additional file 1_CR 150617.docx



1 **Reviewer reports:**2 **Reviewer #2:** Douglas Yu

3 This paper should be published, and the mitogenomes will come in handy. I did, however,
4 check genbank for the assemblies, but they are mostly not available (the two *Catopuma*
5 assemblies are available but under a separate paper). are the assemblies under publication
6 embargo? I have no other comments. [Yes, the data is currently still under publication embargo.](#)

8 **Reviewer #1:** John-James Wilson

9 Thank you for this opportunity to review a GigaScience Data Note.

10 This is an important dataset (as the authors nicely explain) with a vast range of potential
11 usages. In particular, this dataset could be an invaluable resource to the application of mammal
12 monitoring via eDNA and iDNA approaches in Southeast Asia, which is an applied research
13 area in which my lab has done some preliminary work [Genome, 2016, 59(11): 1008-1022,
14 10.1139/gen-2015-0193].

15 The data all appears to be readily available and has been generated following standard
16 protocols from a variety of collections and labs from around the world. As such, this Data Note
17 represents an impressive collaboration. Given that this is largely novel data, any discrepancies
18 in terms of species identifications etc., will only become apparent as more similar data
19 accumulates, and through incorporation of this data into future analyses. Therefore the release
20 of this data is commendable.

21
22 Consequently, my comments are restricted to minor issues regarding the context of the data
23 and the presentation of the Data Note.

25 **Abstract**

26 - The title and abstract highlight Southeast Asia, but samples were obtained from localities as
27 geographically widespread as Sri Lanka (South Asia) and Taiwan (East Asia). While these
28 samples may represent species also found in Southeast Asia (?), it is possible they could be
29 quite divergent (in their mitochondrial sequences) from conspecifics from Southeast Asia
30 proper. In any case, it would be good to justify how the selection of samples obtained for
31 mitogenome sequencing was done, if the authors want to retain the emphasis on this
32 geographic region.

34 [We have removed the three samples below, now making the total mitogenomes 72.](#)

35

No	ID	Voucher	Genebank ID	Genus	Species	Locality
43	MRE	20805.a	KY117561	<i>Muntiacus</i>	<i>reevesi</i>	Shin-Yi, Taiwan
17	54	CN2322	KY117547	<i>Herpestes</i>	<i>hosei</i>	Ceylon
45	MVA	91131	KY052092	<i>Muntiacus</i>	<i>vaginalis</i>	Gangtok, South Sikkim, India

36

37

38

- 39 - Mitochondrion barcodes or mitochondrial barcodes? [Corrected to mitochondrial barcodes](#)
40
41 - The phrase "we have contributed to this need" is a bit confusing. A better phrase may be "we
42 have partially addressed this need".[corrected](#)
43
44 - It may be clearer to say the 32 species had no previous mitogenome available rather than
45 have no current mitogenome available. [corrected](#)
46
47 - It would be good to say how mammal species from SEA now have mitogenomes available.
48 What percentage do the 32 novel mitogenomes from this study contribute?
49 [Currently there are 922 mammalian mitogenomes available in Genbank. Unfortunately most are](#)
50 [not tagged by location/origin. Data mining through manual screening of each mitogenomes](#)
51 [resulted in 174 terrestrial mammal species which are typical to SE Asia. In this work, 30 novel](#)
52 [species are added, contributing to ~17% expansion of the current SE Asia mammal](#)
53 [mitogenome database. You can view the full list in the link below:](#)
54 [https://docs.google.com/spreadsheets/d/1t3pkxCMKbogQjX5YfLGW1waDaD9NL1_eMYzI5CdF](https://docs.google.com/spreadsheets/d/1t3pkxCMKbogQjX5YfLGW1waDaD9NL1_eMYzI5CdFoJo/edit#gid=611706729)
55 [oJo/edit#gid=611706729](https://docs.google.com/spreadsheets/d/1t3pkxCMKbogQjX5YfLGW1waDaD9NL1_eMYzI5CdFoJo/edit#gid=611706729)
56
57 - I'm not sure why Malaysia is a keyword given the broad range of countries from which samples
58 were obtained? [removed](#)
59

60 **Data description**

- 61 -The context is a bit brief, but may be adequate for a Data Note?
62
63 -Could GPS coordinates be given in table 1 and 2? I would find this more useful than the map.
64 [The GPS locations are given in Additional file 2. We decided not to add them in Tables 1 and 2 because](#)
65 [most samples, especially museum specimens don't have their exact location documented \(e.g Sumatra\).](#)
66 [The GPS presents a rough location with accuracy around 100 km. Thus, we thought the map would be](#)
67 [better to show the mitogenome distribution in this work.](#)
68
69
70 -In the tables I suggest not to italicise the vernacular English name, to avoid confusion with
71 scientific names. [corrected](#)
72
73 -In the tables, in the assembly size column, sometimes a comma is used and other times not.
74 [corrected](#)
75
76 -Museum is miss-spelled in the table footers and in other places in the tables. [corrected](#)
77
78 -In my experience West Malaysia is most commonly referred to as Peninsular Malaysia. It would
79 be good to include states and provinces in the localities if known. Also, it would be good to be
80 consistent and always put the country name i.e. Sumatra, Indonesia. Besides, a strange spelling
81 is used for Malacca/Melaka. [corrected](#)
82

83 -Central Siam is present-day Thailand? Ceylon is Sri Lanka? In general it would be good to
84 convert all country/state/city names to their currently used forms. The specimen/museum label
85 form of the locality could be given in another column in the table or in parentheses if necessary.

86 [corrected](#)

87

88

89 -For the DNA extraction protocols (which are given very briefly) it would be good to state that
90 these are standard protocols- that they were performed following the manufacturer's guidelines
91 or following given references etc. I know reference are given in a few cases, but not all.

92 [corrected](#)

93

94 -I'm not sure what the phylogenetic analysis contributes to this Data Note.

95 [It shows a clearer taxonomical order distribution of mitogenomes presented in this work.](#)

96

97

98

99

100

101