

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

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

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

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

7
8 ¹Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Øster
9 Voldgade 5-7, 1350, Copenhagen, Denmark

10 ²Faculty of Biosciences and Medical Engineering, Universiti Teknologi Malaysia, 81310 Johor Bahru,
11 Johor, Malaysia

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

13 ⁴Undergraduate Program on Genomic Sciences, Universidad Nacional Autonoma de Mexico, 62210
14 Cuernavaca, Mexico

15 ⁵BGI-Shenzhen, Shenzhen, GuangDong, China

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

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

18 ⁸University of Potsdam, Institute for Biochemistry and Biology, Karl-Liebknecht-Str 24-25, 14476
19 Potsdam, Germany

20 ⁹Gene Bank of Primates and Primate Genetics Laboratory, German Primate Center, Leibniz Institute for
21 Primate Research, Kellnerweg 4, 37077, Göttingen, Germany

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

23 ¹¹Department of Natural History, Royal Ontario Museum, Toronto, Canada

24 ¹²School of Biological and Chemical Sciences, Queen Mary University of London, Mile End Road, London
25 E1 4NS, United Kingdom

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

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

29 Abstract

30 Background:

31 Southeast (SE) Asia is one of the most biodiverse regions in the world and it holds approximately 20% of
32 all mammal species. Despite this, the majority of SE Asia's genetic diversity is still poorly characterised.
33 The growing interest in using environmental DNA (eDNA) to assess and monitor SE Asian species, in
34 particular threatened mammals - has created the urgent need to expand the available reference database
35 of mitochondrial barcode and complete mitogenome sequences. We have partially addressed this need
36 by generating 72 new mitogenome sequences reconstructed from DNA isolated from a range of historical
37 and modern tissue samples.

38 Findings:

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

44 Conclusion:

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

51 Keywords:

52 invertebrate-derived (iDNA), metabarcoding, GenBank, Taxonomic assignment

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

57 Data description:

58 Context

59 Southeast (SE) Asia is one of the most biodiverse regions in the world, hosting ~20% of mammal species,
60 but is experiencing rapid deforestation for agriculture and development. To assess the ecological
61 consequences of land-use change, there is growing interest in using environmental DNA (eDNA) to
62 monitor mammal populations, particularly threatened taxa that often underpin conservation policies [1–4].
63 Yet current efforts are hampered by the lack of a reference database of mitochondrial barcodes and
64 complete mitogenome sequences. Currently there are 922 mammalian mitogenomes available in
65 Genbank. Unfortunately most are not tagged by location/origin. Data mining through manual screening of
66 each mitogenomes resulted in 174 terrestrial mammal species which are typical to SE Asia. In this work,
67 30 novel species are added, contributing to ~17% expansion of the current SE Asia mammal mitogenome
68 database.

70 DNA extraction

71 Genomic DNA was extracted from different sample types of 72 small mammals, comprising 52 species,
72 listed in Table 1 and Table 2. DNA from modern tissue and blood samples was isolated using the Qiagen
73 DNeasy extraction kit (Qiagen, Hilden, Germany) or Invitex DNA extraction kit (Invitex GmbH, Berlin,
74 Germany), as per standard protocols following the manufacturer's guidelines. Historical samples obtained
75 from the Zoological Museum, Natural History Museum of Denmark, University of Copenhagen (ZM, KU)
76 were treated differently according to type of tissue (Additional file 1a), while at the German Primate
77 Center DNA extraction from museum specimens followed Liedigk et al. (2015) [5]. Complete details of
78 sample information are provided in Additional file 2.

80 Mitogenome sequencing, assembly and annotation

81 Mitogenomes were generated using several approaches. In Copenhagen, author FMS constructed
82 Illumina shotgun libraries with insert sizes ranging between 50 bp to 400 bp. To construct libraries, DNA
83 was sheared to the target size range using Bioruptor® XL (Diagenode, USA) and converted into an
84 Illumina-compatible sequencing library using the NEBNext E6070 kit (New England Biolabs, UK). The

1
2
3
4 85 libraries were PCR amplified with index primers and purified using Qiaquick columns (Qiagen, Hilden,
5
6 86 Germany) according to the manufacturer's instruction (Additional file 1b). Multiple libraries were combined
7
8 87 together into three pools, normalized to 10 nM and sequenced across three lanes of Illumina HiSeq 2500
9
10 88 using SR100bp chemistry. In Berlin and Goettingen, mitogenomes were generated by author PRP and
11
12 89 CR using overlapping PCR products using long range PCR (Additional file 1c) followed by library
13
14 90 construction and MiSeq sequencing, or Sanger sequencing as described in [6] and [5, 7, 8], respectively.
15
16 91 Author RM's mitogenomes were done using methods outlined in [9].
17
18 92
19
20 93 Raw reads for FMS samples were assembled independently by authors FMS and FP using two different
21
22 94 approaches, then compared for consistency. Author FMS trimmed the reads for sequencing adapters, low
23
24 95 quality stretches and leading/tailing Ns using AdapterRemoval 1.2 [10]. The mitochondrial genome was
25
26 96 reconstructed with MITObim v1.8 [11] using the reference mitogenome of the closest species available in
27
28 97 GenBank as the seed reference (Additional file 2). In order to obtain the mapping statistics of the
29
30 98 samples, we ran PALEOMIX v 1.2.6 [12] with default parameters where reads shorter than 25 bp after
31
32 99 trimming were discarded. The trimmed reads were aligned against the newly assembled mitogenome
33
34 100 generated by MITObim using Burrows–Wheeler Aligner (BWA) [13]. Alignments showing low-quality
35
36 101 scores and PCR duplicates were further removed using the MarkDuplicates program from Picard tools,
37
38 102 and reads were locally realigned around small insertions and deletions (indels) to improve overall genome
39
40 103 quality using the IndelRealigner tool from the Genome Analysis Toolkit (GATK)[14]. In contrast, author FP
41
42 104 inputted the trimmed reads into mitoMaker (<https://sourceforge.net/projects/mitomaker/>), which performs a
43
44 105 de-novo and reference-based assembly using SOAPdenovoTrans v1.03 [15] and MITObim v1.7 [11].
45
46 106 Post assembly, the FMS and FP mitogenomes were manually compared for consistency by FMS to
47
48 107 generate the final consensus sequences. These assemblies were automatically annotated using
49
50 108 tRNAscan-SE v1.4 [16] and BLAST v2.2.29 [17] using the mitochondrial genomes found in the NCBI
51
52 109 RefSeq (<https://www.ncbi.nlm.nih.gov/refseq/>) as references.
53
54
55 110
56
57 111 For mitogenome constructed by author RM, Illumina sequence reads were de-multiplexed according to
58
59 112 the respective indexes with the Illumina software bcl2fastq v2.17 (Illumina, San Diego, CA, USA) and
60
61
62
63
64
65

1
2
3
4 113 adapters were clipped from the sequence reads with the software cutadapt v1.3 [18]. Quality trimming
5
6 114 was done through a sliding window approach (10 bp; Q20) and all reads shorter than 20 bp were
7
8 115 removed from the analyses. Mitogenome references from target or closely related species were used for
9
10 116 mapping of the sequencing reads. Aligned reads were de-duplicated using MarkDuplicates from Picard-
11
12 117 tools v1.106 (<https://github.com/broadinstitute/picard>). VariantCalling was carried out using Samtools v1.1
13
14 118 [13] and Bcftools v1.2 (<http://github.com/samtools/bcftools>). For each sample, GATK [14] Variant Calling
15
16
17 119 output files were further filtered to have a minimum read coverage $\geq 3x$, and variants were only called
18
19
20
21 120 when the corresponding base was represented by $\geq 50\%$, otherwise this position was "N"-masked.
22

23 121
24
25 122 Numbers of raw reads generated for each samples and mapping statistics for all 72 mitogenome
26
27 123 assemblies are shown in Additional file 2. Sanger sequenced mitogenomes were checked with 4Peaks
28
29 124 1.8 (www.nucleobytes.com), assembled with SeaView 4.5.4 [19] and annotated with DOGMA [20]. All
30
31 125 mitogenomes were checked manually by eye to identify possible errors caused by insertion and deletions
32
33 126 in Tablet [21]. The final mitochondrial genomes have been uploaded to GenBank (accession numbers are
34
35 127 provided in Table 1 and Table 2). The details of all new mitogenomes assembled in this work are given in
36
37 128 Table 1 and Table 2. Mitogenomes (60 samples) with known localities were geotagged and mapped to
38
39 129 display its geographical distribution (Figure 1).
40

41 130

44 131 **Phylogenetic analysis**

45 132 All the sequenced mitogenomes were aligned using MAFFT v7.158b [22], using the E-INS-i option
46
47 133 (Additional file 3). RAxML v8.0.26 [23] was used to perform the phylogenetic analysis with a
48
49 134 GTR+GAMMA model of nucleotide substitution. To obtain node support, we used 100 bootstrap pseudo-
50
51 135 replicates (Figure 2). The newick file is provided as Additional file 4.
52

53 136

56 137 **Availability of supporting data**

57
58 138 Raw shotgun data are deposited in the SRA under bioproject number PRJNA361218 and is available in the
59
60 139 GigaScience repository, GigaDB <ftp://user27:SallehFaezah@climb.genomics.cn>.
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

140 Abbreviations

141 SE: South East; eDNA: environmental deoxyribonucleic acid; bp: basepair; PCR: polymerase chain
142 reaction; BWA: Burrows–Wheeler Aligner; GATK: Genome Analysis Toolkit; BLAST: Basic Local
143 Alignment Search Tool; NCBI RefSeq: National Center for Biotechnology Information Reference
144 Sequence Database; MAFFT: Multiple Alignment using Fast Fourier Transform ; RAxML: Randomized
145 Axelerated Maximum Likelihood.

146

147

148 Acknowledgements

149 This project was funded by the Malaysian Government, Lundbeck Foundation (R52-5062), Leibniz-
150 Association grant SAW-2013-IZW-2, the German Federal Ministry of Education and Research grant
151 BMBF FKZ: 01LN1301A and the German Primate Center. We thank the Danish National High-throughput
152 Sequencing Centre for assistance in generating Illumina data. We also thank the staff of the zoos in
153 Hannover, Copenhagen, Gettorf, Belfast, Bristol, Howletts Wild Animal Park, the staff of the Ludwig-
154 Maximilians-University Munich, the Bavarian State Collection Munich, the National Museums Scotland,
155 Edinburgh, the Natural History Museum Berlin, Naturalis Leiden, and the Zoological Museum, Natural
156 History Museum of Denmark, University of Copenhagen for providing samples for this work. BKL
157 gratefully acknowledges the Indonesian and Malaysian government authorities for issuing scientific
158 permits.

159

160 Authors' contributions

161 FMS, AW, JF and MTPG conceived the project. FMS, MHSS, MSS, MSA, RM, PRP, CR, BKL, and SJR
162 collected the samples and extracted the genomic DNA. FMS, RM, PRP and CR constructed the libraries
163 and did sequencing. FMS, JRM, FP, SL, PRP, RM, DL and CR assembled the mitogenomes and
164 performed mitogenome analysis. FMS, SL, PRP and MTPG wrote the article. All authors discussed the
165 project and data. All authors read and approved the final manuscript.

166

167

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

168 Competing interests

169 The authors declare that they have no competing interests.

170

171 Author details

172 ¹Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Øster

173 Voldgade 5-7, 1350, Copenhagen, Denmark. ²Faculty of Biosciences and Medical Engineering, Universiti

174 Teknologi Malaysia, 81310 Johor Bahru, Johor, Malaysia. ³Leibniz Institute for Zoo and Wildlife Research,

175 Alfred-Kowalke Strasse 17, 10315 Berlin, Germany. ⁴Undergraduate Program on Genomic Sciences,

176 Universidad Nacional Autonoma de Mexico, 62210 Cuernavaca, Mexico. ⁵BGI-Shenzhen, Shenzhen,

177 GuangDong, China. ⁶Natural History Museum, University of Oslo, P.O. Box 1172 Blindern, NO-0318,

178 Oslo, Norway. ⁷Freie Universität Berlin, Kaiserswerther Str. 16-18, 14195 Berlin, Germany. ⁸University of

179 Potsdam, Institute for Biochemistry and Biology, Karl-Liebknecht-Str 24-25, 14476 Potsdam, Germany.

180 ⁹Gene Bank of Primates and Primate Genetics Laboratory, German Primate Center, Leibniz Institute for

181 Primate Research, Kellnerweg 4, 37077, Göttingen, Germany. ¹⁰Forest Biodiversity Division, Forest

182 Research Institute Malaysia, 52109 Kepong, Selangor, Malaysia. ¹¹Department of Natural History, Royal

183 Ontario Museum, Toronto, Canada. ¹²School of Biological and Chemical Sciences, Queen Mary

184 University of London, Mile End Road, London E1 4NS, United Kingdom. ¹³NTNU University Museum,

185 Norwegian University of Science and Technology, Trondheim, Norway.

186

187

188

189

190

191

192

193

194

195

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

References

- 196 1. Bohmann K, Evans A, Gilbert MTP, Carvalho GR, Creer S, Knapp M, Yu DW, de Bruyn M (2014)
197 Environmental DNA for wildlife biology and biodiversity monitoring. *Trends Ecol Evol* 29:358–367
198
- 199 2. Lee P-S, Gan HM, Clements GR, Wilson J-J (2016) Field calibration of blowfly-derived DNA against
200 traditional methods for assessing mammal diversity in tropical forests. *Genome* 59:1008–1022
201
- 202 3. Schnell IB, Sollmann R, Calvignac-Spencer S, Siddall ME, Douglas WY, Wilting A, Gilbert MTP
203 (2015) iDNA from terrestrial haematophagous leeches as a wildlife surveying and monitoring tool--
204 prospects, pitfalls and avenues to be developed. *Front Zool* 12:24
205
- 206 4. Schnell IB, Thomsen PF, Wilkinson N, Rasmussen M, Jensen LRD, Willerslev E, Bertelsen MF,
207 Gilbert MTP (2012) Screening mammal biodiversity using DNA from leeches. *Curr Biol* 22:R262–3
208
- 209 5. Liedigk R, Kolleck J, Böker KO, et al (2015) Mitogenomic phylogeny of the common long-tailed
210 macaque (*Macaca fascicularis fascicularis*). *BMC Genomics* 16:222
211
- 212 6. Patel RP, Förster DW, Kitchener AC, et al (2016) Two species of Southeast Asian cats in the genus
213 *Catopuma* with diverging histories: an island endemic forest specialist and a widespread habitat
214 generalist. *Open Science* 3:160350
215
- 216 7. Roos C, Zinner D, Kubatko LS, et al (2011) Nuclear versus mitochondrial DNA: evidence for
217 hybridization in colobine monkeys. *BMC Evol Biol* 11:77
218
- 219 8. Liedigk R, Roos C, Brameier M, Zinner D (2014) Mitogenomics of the Old World monkey tribe
220 Papionini. *BMC Evol Biol* 14:176
221
- 222 9. Fortes GG, Paijmans JLA (2015) Analysis of Whole Mitogenomes from Ancient Samples. *Methods*
223 *Mol Biol* 1347:179–195
224
- 225 10. Lindgreen S (2012) AdapterRemoval: easy cleaning of next-generation sequencing reads. *BMC Res*
226 *Notes* 5:337
227
- 228 11. Hahn C, Bachmann L, Chevreur B (2013) Reconstructing mitochondrial genomes directly from
229 genomic next-generation sequencing reads—a baiting and iterative mapping approach. *Nucleic*
230 *Acids Res.*
231
- 232 12. Schubert M, Ermini L, Der Sarkissian C, et al (2014) Characterization of ancient and modern
233 genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX. *Nat*

1
2
3
4 224 Protoc 9:1056–1082
5
6 225 13. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows–Wheeler transform.
7
8 226 Bioinformatics 25:1754–1760
9
10 227 14. McKenna A, Hanna M, Banks E, et al (2010) The Genome Analysis Toolkit: a MapReduce framework
11
12 228 for analyzing next-generation DNA sequencing data. *Genome Res* 20:1297–1303
13
14 229 15. Xie Y, Wu G, Tang J, et al (2014) SOAPdenovo-Trans: de novo transcriptome assembly with short
15
16 230 RNA-Seq reads. *Bioinformatics* 30:1660–1666
17
18 231 16. Lowe TM, Eddy SR (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in
19
20 232 genomic sequence. *Nucleic Acids Res* 25:955–964
21
22 233 17. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL (2009) BLAST+:
23
24 234 architecture and applications. *BMC Bioinformatics* 10:421
25
26 235 18. Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads.
27
28 236 *EMBnet.journal* 17:10–12
29
30 237 19. Gouy M, Guindon S, Gascuel O (2010) SeaView version 4: A multiplatform graphical user interface
31
32 238 for sequence alignment and phylogenetic tree building. *Mol Biol Evol* 27:221–224
33
34 239 20. Wyman SK, Jansen RK, Boore JL (2004) Automatic annotation of organellar genomes with DOGMA.
35
36 240 *Bioinformatics* 20:3252–3255
37
38 241 21. Milne I, Stephen G, Bayer M, Cock PJA, Pritchard L, Cardle L, Shaw PD, Marshall D (2012) Using
39
40 242 Tablet for visual exploration of second-generation sequencing data. *Brief Bioinform* bbs012
41
42 243 22. Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements
43
44 244 in performance and usability. *Mol Biol Evol* 30:772–780
45
46 245 23. Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large
47
48 246 phylogenies. *Bioinformatics* 30:1312–1313
49
50
51 247
52
53 248
54
55 249
56
57 250
58
59 251
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

252 Tables and Figures

- 253 1. Table 1: List of mitogenomes assembled in this work that supplement pre-existing mitogenome
- 254 references already available in GenBank
- 255 2. Table 2: List of mitogenomes assembled in this work that have no previous complete mitogenome
- 256 reference available in GenBank
- 257 3. Figure 1: Geographical distribution of mitogenomes assembled in this work (60 mitogenomes with
- 258 known locality)
- 259 4. Figure 2: Phylogenetic tree of mitogenomes assembled in this work

260

261

262 Additional files

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

267

268

269

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	Copenhagen Zoo	ZM, KU	21/8/1912	FMS/FP
3	KX265095	Bay Cat	<i>Catopuma</i>	<i>badia</i>	16960	Sabah, Malaysia	National Museum Scotland	20/04/2000	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
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

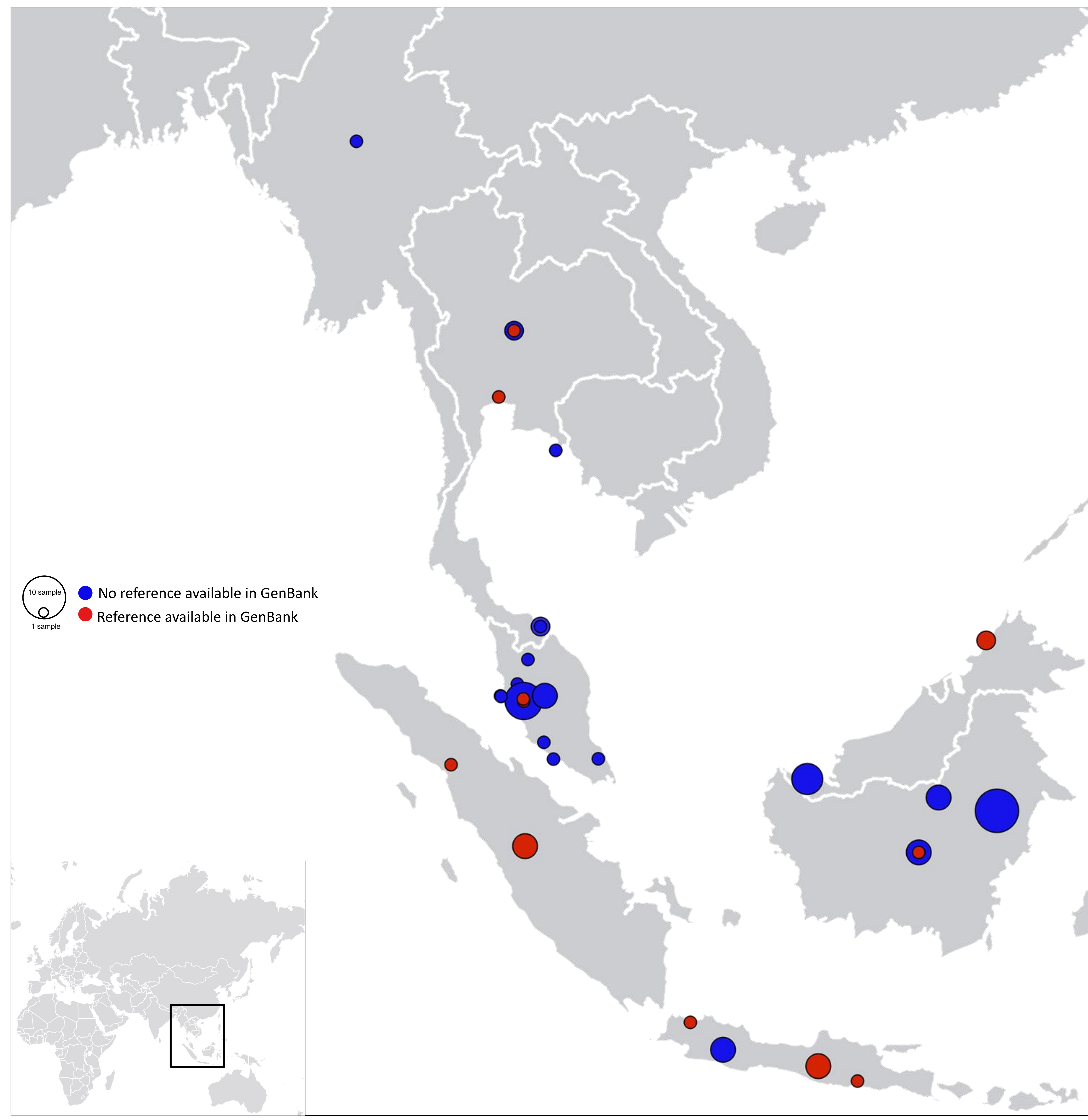
Table 2: List of mitogenomes assembled in this work that have no previous complete mitogenome reference available in GenBank

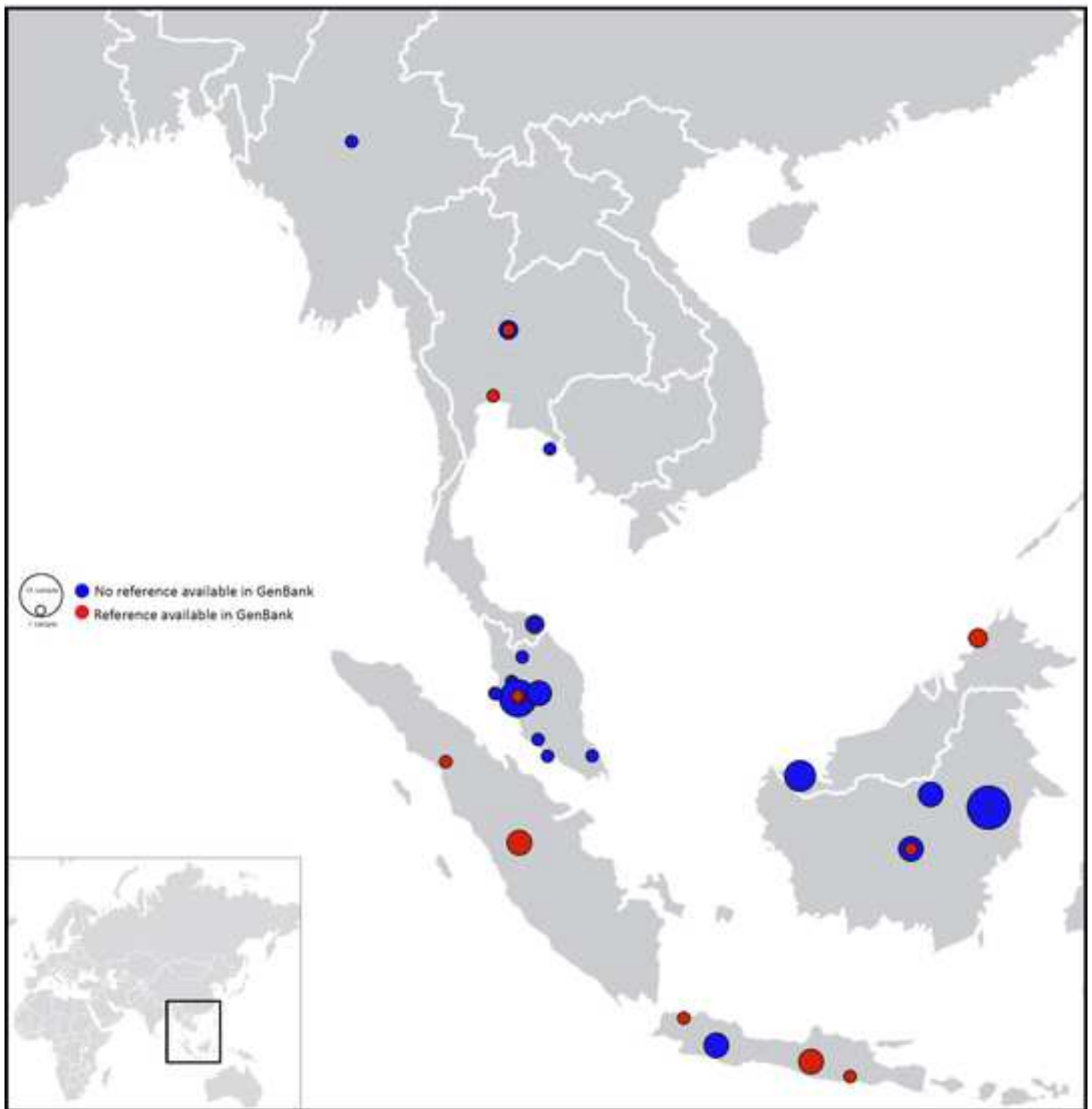
No	GenBank ID	Common name	Genus	Species	Assembly Size	Locality	Source	Sample date of collection	Data by
1	KY117536	Asian Small-clawed Otter	<i>Aonyx</i>	<i>cinereus</i>	16153	captive	Copenhagen Zoo	08/08/11	FMS/FP
2	KY117535	Asian Small-clawed Otter	<i>Aonyx</i>	<i>cinereus</i>	16153	Sarawak, Malaysia	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</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	10.11.1931	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	name pending	<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

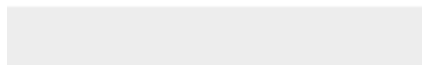
ZM, KU: Zoological Museum, University of Copenhagen; ROM: Royal Ontario Museum; FRIM: Forest Research Institute, Malaysia; NHM: Natural History Museum

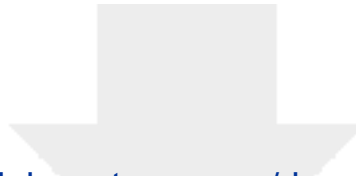






Click here to access/download
Supplementary Material
READMEsallehetal.txt

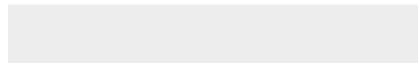


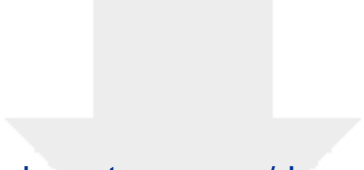


Click here to access/download

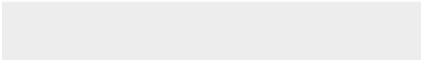

Supplementary Material

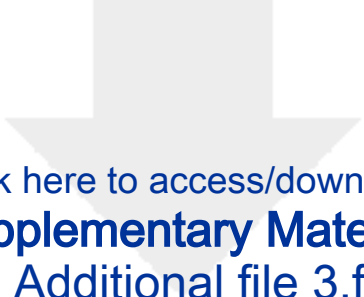
1DatainformationSallehetal.docx



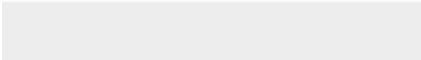



Click here to access/download
Supplementary Material
5 Additional file 2.xlsx



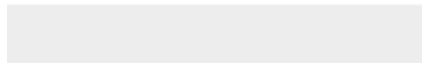


Click here to access/download
Supplementary Material
5 Additional file 3.fa



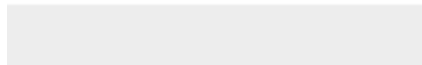


Click here to access/download
Supplementary Material
5 Additional file 4 newick.txt





Click here to access/download
Supplementary Material
5Additional file 1.docx





Click here to access/download
Supplementary Material
Correctionsforsubmissions.docx

