Data in brief 25 (2019) 104133

Contents lists available at ScienceDirect

Data in brief

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

Data Article

Partial mtDNA sequencing data of vulnerable *Cephalopachus bancanus* from the Malaysian Borneo

Muhamad Aidil Zahidin [a, b, *], Norehan Abd Jalil [b, c], Nur Mukminah Naharuddin [b], Mohd Ridwan Abd Rahman [d], Millawati Gani [b, e], Mohd Tajuddin Abdullah [a, f, **]

[a] *Institute of Tropical Biodiversity and Sustainable Development, Universiti Malaysia Terengganu, 21030 Kuala Nerus, Malaysia*
[b] *Faculty of Resource Science and Technology, Universiti Malaysia Sarawak, 94300 Kota Samarahan, Malaysia*
[c] *Faculty of Science, Universiti Putra Malaysia, 43400 UPM Serdang, Malaysia*
[d] *Center for Pre-University Study, Universiti Malaysia Sarawak, 94300 Kota Samarahan, Malaysia*
[e] *National Wildlife Forensic Laboratory, Department of Wildlife and National Park Peninsular Malaysia, KM 10, Jalan Cheras, 56100 Kuala Lumpur, Malaysia*
[f] *School of Marine and Environmental Sciences, Universiti Malaysia Terengganu, 21030 Kuala Nerus, Malaysia*

ARTICLE INFO

Article history:
Received 29 January 2019
Received in revised form 25 May 2019
Accepted 3 June 2019
Available online 25 June 2019

Keywords:
Cytochrome b gene
Population genetics
Primate
Tarsier
Borneo

ABSTRACT

Tarsier is an endangered nocturnal primate in the family Tarsiidae and is an endemic to Sundaic islands of Philippine (*Carlito syrichta*), Sulawesi (*Tarsius* tarsier-complex) and Borneo (*Cephalopachus bancanus*). Recent records indicated that most molecular studies were done on the Eastern Tarsier and little information for the other group of tarsiers. Here, we present a partial cytochrome b data set of *C. bancanus* in Sarawak, Malaysian Borneo. Standard mist nets were deployed at strategic locations in various habitat types. A total of 18 individuals were caught, measured and weighed. Approximately, 2 × 2 mm of tissue samples were taken and preserved in molecular grade alcohol. Out of 18, only 11 samples were screened with partial mtDNA (cytochrome *b*) and the DNA sequences were registered in the GenBank (accession

* Corresponding author. Institute of Tropical Biodiversity and Sustainable Development, Universiti Malaysia Terengganu, 21030 Kuala Nerus, Malaysia.
** Corresponding author. Institute of Tropical Biodiversity and Sustainable Development, Universiti Malaysia Terengganu, 21030 Kuala Nerus, Malaysia.
E-mail addresses: aiditarsier@gmail.com (M.A. Zahidin), mohd.tajuddin@umt.edu.my (M.T. Abdullah).

https://doi.org/10.1016/j.dib.2019.104133

numbers: KY794797-KY794807). Phylogenetic trees were constructed with 20 additional mtDNA sequences downloaded from GenBank. The data are valuable for the management authorities to regulate the type of management units for the metapopulation to sustain population genetics integrity of tarsiers in the range countries across the Sunda Shelf.

Specifications table

Subject area	Biology
More specific subject area	Molecular Evolution
Type of data	Cytochrome *b* partial data are presented as in Tables 1–6, Figs. 1–3 and Supplementary Tables 1–3.
How data was acquired	Data were acquired by extracting and amplifying, purifying (Promega Wizard SV Gel and PCR Clean-Up System (Promega Co.), and sequencing (First Base Laboratories Malaysia) the target mtDNA region and analysed using Sequencher 5.4 (https://genecodes.com), ClustalW2 MUSCLE (https://www.ebi.ac.uk), MEGA 7 and DnaSP software.
Data format	Raw and analysed data
Experimental factors	The sequence alignments were trimmed and filtered
Experimental features	Phylogenetic analyses of partial cytochrome *b*
Data source location	Sarawak, Malaysian Borneo and GenBank
Data accessibility	GenBank with accession number KY794797-KY794807 (https://www.ncbi.nlm.nih.gov/nuccore/?term=Cephalopachus+bancanus+bancanus+isolate)
Related research article	M.T. Abdullah, Mammalian Evolution and Biogeography (Evolusi dan Biogeografi Mammalia), Universiti Malaysia Terengganu, Kuala Nerus 2016.

Value of the Data

- The data are valuable for the management authorities to determine the type of management units for the metapopulations to maintain the integrity of population genetics in their ranges across the Sunda Shelf.
- The data can be used as baseline information for future studies on genetic and molecular ecology that can be used as a flagship model to test the "Out of Sunda" theory and elucidating the history of prehistoric humans and primates migration waves in Southeast Asia.
- The data allow other researchers focusing on this population to start the genome-wide analysis.

1. Data

Tarsiers are a vulnerable primate group [1] in family Tarsiidae that can be found on Southeast Asia Islands; Sundaic islands of Philippine (*Carlito syrichta*), Sulawesi and surrounding islands (*Tarsius* tarsier-complex) and Borneo (*Cephalopachus bancanus*) [2]. Western Tarsier *Cephalopachus bancanus bancanus* can be found in Malaysian Borneo and is listed as protected and totally protected species in the Malaysia's Wildlife Conservation Act (WCA) 2010 and Sarawak's Wildlife Protection Ordinance (WLPO) 1998 respectively. The molecular research interest on this endemic species is due to the availability of recent information related to taxonomy and evolutionary relationship of tarsiers since the expansion of fauna and prehistoric human into Southeast Asia [2,3].

This dataset contains genetic phylogenetic information of *C. bancanus* from Malaysian Borneo. Table 1 shows a list of field sampling conducted in Sarawak, Borneo. Field number, standard morphological measurements, weight and sex of each individual were recorded as in Table 2. A set of partial primers of Cytochrome *b*, DNA master mixture profile and PCR profile were tabulated as in Table 3 and Supplementary Tables 1 and 2 respectively [4]. Additional 20 mtDNA sequences derived from the GenBank [5–15] were used and tabulated in Table 4. The sequence variations, frequency distribution haplotypes and pairwise distance of tarsier were identified as in Tables 5 and 6 and Supplementary

Table 1
Field sampling conducted in Sarawak, Borneo.

	Division	Sampling site	Coordinate
1	Betong	Maludam National Park	1.5271° N, 111.1414° E
2	Kota Samarahan	Universiti Malaysia Sarawak	1.4649° N, 110.4269° E
3	Kuching City	Cermat Ceria Forest	1.° 24′01.6″ N, 111° 23′54.0″ E
4	Kuching City	Durafarm Plantaion	1° 23′50.63697″ N, 111° 50.59624″ E
5	Kuching City	Kampung Barieng	1° 25′0″ N, 110° 0′9″ E
6	Kuching City	Kubah National Park	1.6128° N, 110.1969° E
7	Kuching City	Matang Wildlife Centre	1.6166° N, 110.1582° E

Table 2
Taxonomic measurements of captured *C. bancanus* with their registered accession number in the GenBank.

	Field no.	Species	Measurements (mm)										Wt (g)	Sex	Note	Accession Number
			E	HF	TV	HB	TL	RH	LH	RF	LF	Chest				
1	TSKNP001	*Cephalopachus bancanus*	40.23		214	130	344	48.04	46.96	42.95	42.53	110	68	M	Kubah National Park	
2	TSKNP002	*C. bancanus*	27.71	40.79	206	136	342	46.03	45.67	40.79	40.71	126.93	150	F	Kubah National Park	
3	TSC002	*C. bancanus*													UNIMAS	
4	TSC003	*C. bancanus*	27.60		191	149	340	40.64	40.32	36.9	37.3		105	F	UNIMAS	KY794803
5	TSC004	*C. bancanus*	38.66		225	143	368	49.3	50	39.1	39		110	F	UNIMAS	KY794804
6	TSC005	*C. bancanus*	28.00	65.64	24.6	117	423	45	45	35	35	67		M	UNIMAS	KY794805
7	TSMW001	*C. bancanus*	30.00		138			45	47	40	40	115	110	M	Matang Wildlife Centre	
8	TSMW002	*C. bancanus*	28.00	67.00	216	119		47	47	37	37	92		M	Matang Wildlife Centre	KY794807
9	MNP001	*C. bancanus*	23.00	72.00	241								121	M	Maludam National Park	
10	MNP002	*C. bancanus*	21.80	73.99	200								124	M	Maludam National Park	KY794806
11	PSF001	*C. bancanus*	31.00		266	140	406						115	M	Cermat Ceria Forest	KY794801
12	PSF002	*C. bancanus*	25.23		220	154	374						130	M	Cermat Ceria Forest	
13	KBSM1302	*C. bancanus*	31.00	76.00	225	132	357						119	F	Kampung Barieng	KY794797
14	KBSM1303	*C. bancanus*	22.00	71.00	225	150	375						125	M	Kampung Barieng	KY794798
15	KBSM1304	*C. bancanus*	30.00	70.00	219	140	359						108	F	Kampung Barieng	KY794799
16	KBSM1305	*C. bancanus*	25.00	74.00	225	150	375						123	M	Kampung Barieng	KY794800
17	A08897	*C. bancanus*	21.07	64.62	210	133	343						133	M	Durafarm Plantation	
18	A11251	*C. bancanus*	20.05	76.00	230	141	371							M	Durafarm Plantation	KY794802

E- Ear length, HF- Hind foot length, T- Tail length, HB- Height body length, TL- Total length, RH- Right hand length, LH- Left hand length, RF- Right foot length, LF- Left foot length.
M- Male, F- Female, UNIMAS- Universiti Malaysia Sarawak.

Table 3
Primer used for PCR amplification [4].

Primer	Primer sequences (5′-3′)	Size (bp)
Glud-GL (F)	5′- TGACCTGARAACCAYCGTTG -3′	500
CB2H (R)	5′- CCTTCAGAATGATATTTGTCCTCA -3′	500

Table 4
Additional 20 mtDNA sequences used in this study.

	Scientific name	Common name	Accession Number	Author
1	*Cephalopachus bancanus*	Western tarsier	NC002811	[5]
2	*C. bancanus*	Western tarsier	AF348159	[5]
3	*C. bancanus*	Western tarsier	AB011077	[6]
4	*Carlito syrichta*	Philippine's tarsier	AB371090	[7]
5	*C. syrichta*	Philippine's tarsier	NC012774	[7]
6	*Tarsius wallacei*	Eastern tarsier	HM115983	[8]
7	*T. wallacei*	Eastern tarsier	HM115984	[8]
8	*T. wallacei*	Eastern tarsier	HM115982	[8]
9	*T. lariang*	Eastern tarsier	FJ614357	[9]
10	*T. lariang*	Eastern tarsier	FJ614358	[9]
11	*T. lariang*	Eastern tarsier	FJ614363	[9]
12	*T. dentatus*	Eastern tarsier	FJ614369	[9]
13	*T. dentatus*	Eastern tarsier	FJ614370	[9]
14	*T. dentatus*	Eastern tarsier	FJ614371	[9]
15	*Hylobates muelleri*	Bornean gibbon	Y13300	[10]
16	*Macaca fascicularis*	Long-tailed macaque	AF295584	[11]
17	*Trachypithecus cristatus*	Silvered-leaf monkey	NC023971	[12]
18	*Nasalis larvatus*	Proboscis monkey	DQ355298	[13]
19	*Presbytis hosei*	Hose's langur	JF295114	[14]
20	*Tupaia glis*	Common treeshrew	AY321644	[15]

Table 5
Sequence variation of Western Tarsier.

Indices	Partial Cyt *b*
Base pair	375 bp
Conserved site	366
Variable site	9
Parsimony-informative site	5
Singleton	4
Nucleotide composition (%) C	26.40
T	30.20
A	27.20
G	16.20
Overall mean distance	0.007

Table 6
Frequency distribution of the partial Cyt *b* haplotypes.

Haplotype	n	Sample	Frequency
Hap 1	1	*C. bancanus* TSC003	0.091
Hap 2	1	*C. bancanus* TSC004	0.091
Hap 3	3	*C. bancanus* TSC005, *C. bancanus* KBSM0213, *C. bancanus* A11261	0.273
Hap 4	2	*C. bancanus* TSMW002, *C. bancanus* PSF001	0.182
Hap 5	1	*C. bancanus* MNP002	0.091
Hap 6	2	*C. bancanus* KBSM0313, *C.bancanus* KBSM0513	0.182
Hap 7	1	*C. bancanus* KBSM0413	0.091

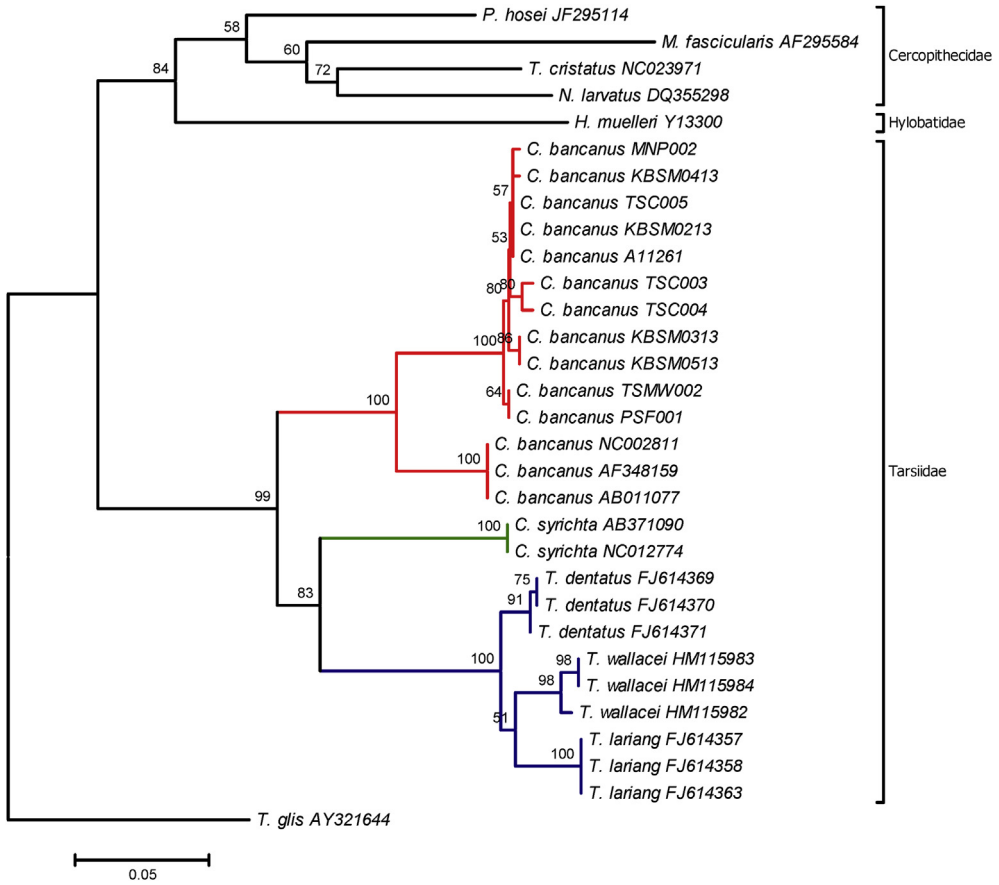

Fig. 1. The evolutionary history was inferred using the Neighbor-Joining tree method. The optimal tree with the sum of branch length = 1.16079630 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1,000 replicates; more than 50%) is shown above the branch. The institutional codes are listed in Tables 2 and 4.

Table 3. The evolutionary relationships of taxa were inferred using the Neighbor-Joining, Maximum Parsimony and Maximum Likelihood methods are shown as in Figs. 1–3.

2. Experimental design, materials and methods

2.1. Sample Collection

Field sampling was conducted at the southern part of Sarawak; Kubah National Park, Matang Wildlife Centre, Universiti Malaysia Sarawak (UNIMAS), Maludam National Park, Cermat Ceria Forest, Kampung Barieng and Durafarm Plantation (Table 1). The samplings were assisted by the field assistants from the Institute of Biodiversity and Environmental Conservation (IBEC), UNIMAS. A total of ten mist nets were deployed at strategic locations with high vegetation, trees with small trunk diameter and near to the stream or water bodies [16–20]. A total of 18 individuals were captured, identified, sexed, measured and weighed (Table 2) [18–21]. Each was tranquilised using Zoletil 100 mg solution. Approximately, 2 × 2 mm-thick tissues samples were taken and preserved in molecular grade alcohol.

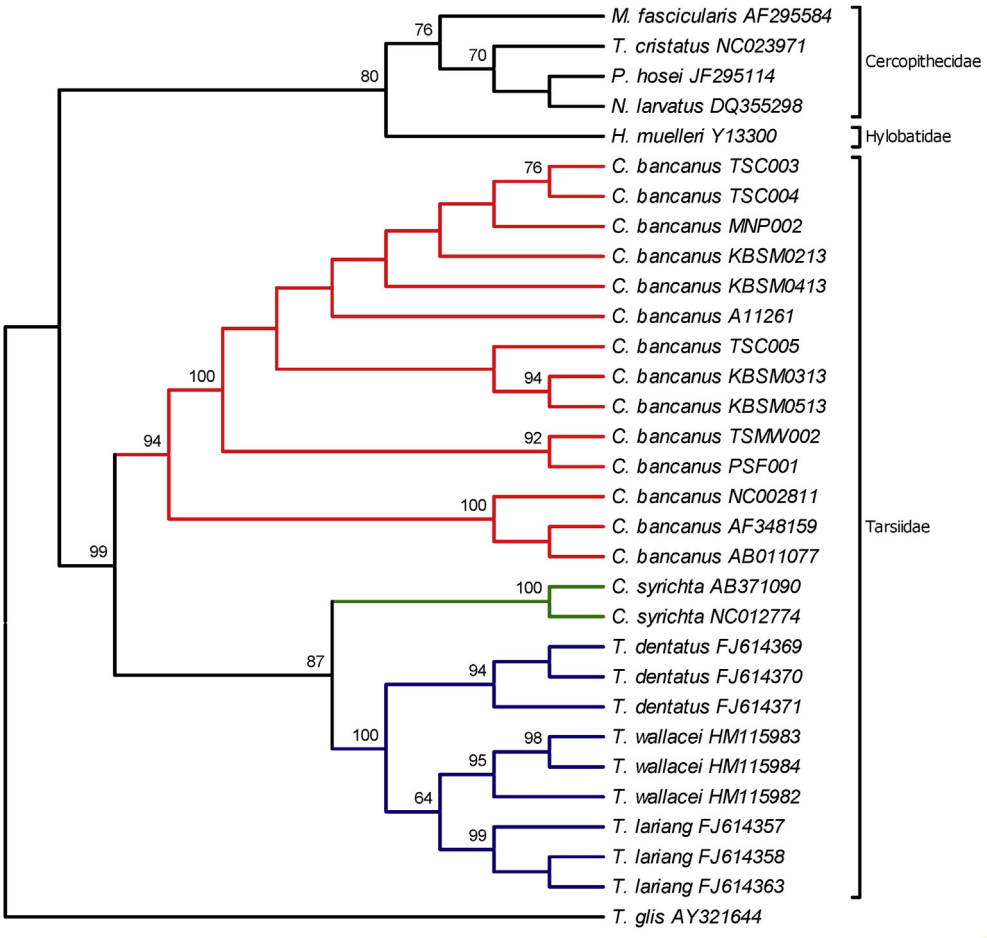

Fig. 2. The evolutionary history was inferred using the Maximum Parsimony method. The tree was obtained using the Subtree-Pruning-Regrafting (SPR) algorithm which the initial trees were obtained by the random addition of sequences. The consistency index is 0.819864 and the composite index is 0.478254 for all sites and parsimony-informative sites. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1,000 replicates; more than 50%) is shown above the branch. The institutional codes are listed in Tables 2 and 4.

2.2. DNA extraction, amplification, purification and sequencing

The DNA samples were extracted using cetyl-tri-methyl ammonium bromide (CTAB) protocol [22] and polymerase chain reaction (PCR) amplified using a set of cytochrome *b* partial primers [4]. The amplified products were purified using Promega Wizard SV Gel and PCR Clean-Up System (Promega Co.) and subjected to cycle sequencing at the First Base Laboratories Malaysia. The *C. bancanus* sequences were registered in the GenBank (accession numbers: KY794797-KY794807) (Table 2).

2.3. Sequence analysis

The nucleotide sequences were visualized and read using Sequencher 5.4 (https://genecodes.com). The sequences were matched and aligned with 20 additional mtDNA sequences (Table 4) [5–15] using ClustalW2 MUSCLE (Multiple Sequence Comparison by Log-Expectation) (https://www.ebi.ac.uk). The

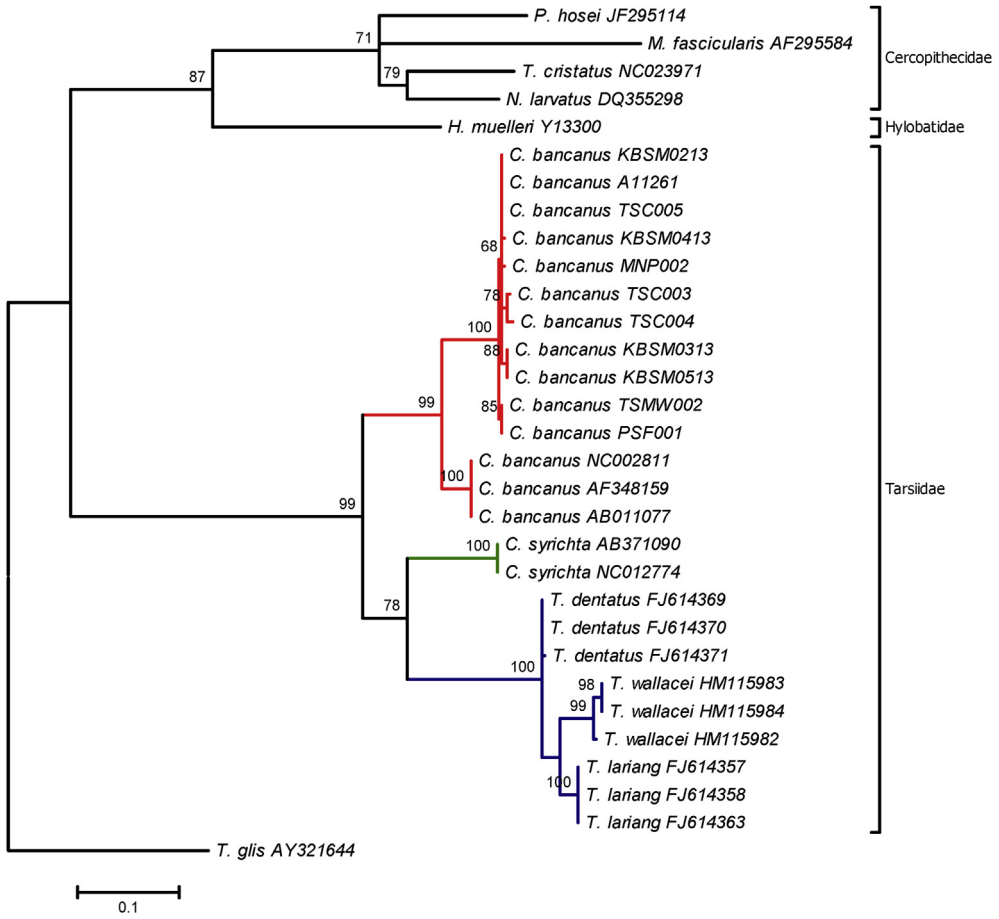

Fig. 3. The evolutionary history was inferred by using the Maximum Likelihood method based on the Hasegawa-Kihino-Yano (HKY + G + I) model and the tree with the highest log likelihood (−2336.6352) is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1,000 replicates; more than 50%) is shown above the branch. The institutional codes are listed in Tables 2 and 4.

nucleotide composition and haplotype frequency were performed in Molecular Evolutionary Genetics Analysis (MEGA) 7 [23] and DnaSP [24]. The evolutionary divergence between sequences (Supplementary Table 3) was estimated in MEGA 7 by using the p-distance model where all positions containing gaps and missing data were eliminated. Kimura 2-parameter method was used to compute the Neighbor-Joining tree (Fig. 1). The evolutionary history of Maximum Parsimony was shown in Fig. 2. The tree was obtained using the Subtree-Pruning-Regrafting (SPR) algorithm which the initial trees were obtained by the random addition of sequences. Meanwhile, the evolutionary history of Maximum Likelihood was performed using the Hasegawa-Kishino-Yano (HKY + G + I) method (Fig. 3). The best model was chosen based on the Akaike Information Criterion (AIC; 4776.487) value and the lowest Bayesian Information Criterion (BIC; 5254.204) score.

Acknowledgments

We thank geneticist Dr Edinur Hisyam Atan from the Forensic Science Programme, Universiti Sains Malaysia (USM) for insightful comments on various versions of this manuscript. We also thank

geneticist Dr Faisal Ali Anwarali Khan, science officer Mr Wahap Marni, postgraduate and undergraduate students of UNIMAS batch 2013/2014 of Department of Zoology, Faculty of Resource Science and Technology (FRST) for their assistance during field samplings. This study was supported by the Sarawak Forest Department (SFD) researches permit (NCCD/907/4/4/Jld9-39) and park permit (81/2013). This study was made possible under grants RAG/S(4)/916/2012(17) awarded to Mr Mohd Ridwan Abd Rahman and colleagues. The Universiti Malaysia Terengganu funded this publication from the Research and Innovation Management Centre (RIMC) publication grant is gratefully acknowledged.

Conflict of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.dib.2019.104133.

References

[1] M. Shekelle, I. Yustian, *Tarsius bancanus*, the IUCN red list of threatened species 2008, e.T21488A9286601, retrieved from https://doi.org/10.2305/IUCN.UK.2008.RLTS.T21488A9286601.en, accessed on 24 January 2019.
[2] M.A. Zahidin, N. Abd Jalil, N.M. Naharuddin, M.R. Abd Rahman, M. Gani, M.T. Abdullah, Sequence variation and phylogenetic of Western Tarsier (Primate: *Cephalopachus bancanus*) inferred using partial sequences of the cytochrome b segment of the mitochondrial DNA, Under review in Springer Nature International Publication (Under review).
[3] M.A. Zahidin, W.B. Wan Omar, W.R. Wan Taib, J.J. Rovie Ryan, M.T. Abdullah, Sequence polymorphism and haplogroup data if the hypervariable regions on mtDNA in Semoq Beri population, Data Brief 21 (2018) 2609–2615.
[4] S. Palumbi, R. Martin, S. Romano, W.O. McMilan, L. Stice, G. Grabowski, The Simple Fool's Guide to PCR, University of Hawaii, Honolulu, 1991.
[5] J. Schmitz, M. Ohme, H. Zischler, The complete mitochondrial sequence of *Tarsius bancanus*: evidence for an extensive nucleotide compositional plasticity of primate mitochondrial DNA, Mol. Biol. Evol. 19 (2002) 544–553.
[6] T.D. Andrews, L.S. Jermiin, S. Easteal, Accelerated evolution of cytochrome *b* in simian primates: adaptive evolution in concert with other mitochondrial proteins? Mol. Evol. 47 (1998) 249–257.
[7] A. Matsui, F. Rakotondraparany, I. Munechika, M. Hasegawa, S. Horai, Molecular phylogeny and evolution of prosimians based on complete sequences of mitochondrial DNAs, Gene 441 (2009) 53–66.
[8] S. Merker, C. Driller, H. Dahruddin, W. Sinaga, D. Perwitasari-Farajallah, M. Shekelle, *Tarsius wallacei*: a new tarsier species from Central Sulawesi occupies a discontinuous range, Int. J. Primatol. 31 (2010) 1107–1122.
[9] S. Merker, C. Driller, D. Perwitasari-Farajallah, J. Pamungkas, H. Zischler, Elucidating geological and biological processes underlying the diversification of Sulawesi tarsiers, Proc. Natl. Acad. Sci. Unit. States Am. 106 (2009) 8459–8464.
[10] L.M. Hall, D.S. Jones, B.A. Wood, Evolution of the Gibbon subgenera Inferred from Cytochrome b DNA sequence data, Mol. Phylogenetics Evol. 10 (1998) 281–286.
[11] R.V. Collura, P.K. Karanth, C.B.R. Stewart, reportA Cytochrome B Phylogeny of the Colobine Monkeys. Unpublished report. Retrieved from https://www.ncbi.nlm.nih.gov/nuccore/AF295584. Assessed on 22 March 2014.
[12] Q.Y. Jiang, Y.N. Huang, reportStudy on the Complete Mitochondrial Genome of Trachypithecus. Unpublished report. Retrieved from https://www.ncbi.nlm.nih.gov/nuccore/NC_023971. Assessed on 22 March 2014.
[13] K.N. Sterner, R.L. Raaum, Y.P. Zhang, C.B. Stewart, T.R. Disotell, Mitochondrial data support an odd-nosed colobine clade, Mol. Phylogenetics Evol. 40 (2006) 1–7.
[14] D. Meyer, I.D. Rinaldi, H. Ramlee, D. Perwitasari-Farajallah, J.K. Hodges, C. Roos, Mitochondrial phylogeny of leaf monkeys (genus *Presbytis*, Eschscholtz, 1821) with implications for taxonomy and conservation, Mol. Phylogenetics Evol. 59 (2011) 311–319.
[15] K.H. Han, F.H. Sheldon, R.B. Stuebing, reportPhylogenetic diversification and taxonomic status of the *Tupaia glis* species complex based on cytochrome-b sequence variation. Unpublished report. Retrieved from https://www.ncbi.nlm.nih.gov/nuccore/AY321644. Assessed on 22 March 2014.
[16] M.T. Abdullah, A Note on the Chance Preying by a Western Tarsier on Fruit Bats in Malaysian Borneo, retrieved from, 1998, https://www.researchgate.net/publication/330988997_A_Note_on_the_Chance_Preying_by_a_Western_Tarsier_on_Fruit_Bats_in_Malaysian_Borneo. (Accessed 24 January 2019).
[17] M.T. Abdullah, Mammalian Evolution and Biogeography (Evolusi Dan Biogeografi Mammalia), Universiti Malaysia Terengganu, Kuala Nerus, 2016.
[18] C. Karim, A.A. Tuen, M.T. Abdullah, Mammals, Sarawak Mus, J. Special Issue 80 (2004) 221–234.
[19] M.A. Zahidin, A. Roslan, W. Marni, M. Kombi, M.T. Abdullah, Consortium of bako researchers, biodiversity checklist of faunal diversity in bako national park, Sarawak, Malaysia, J. Sustain. Sci. Manage. 11 (2016) 53–72.
[20] J.J. Rovie-Ryan, M. Gani, Y.P. Lee, H.M. Gan, M.T. Abdullah, Mitogenome data of *Nycticebus coucang insularis* Robinson, 1917, (Primate: Lorisidae), Data in Brief 25 (2019) 104058, https://doi.org/10.1016/j.dib.2019.104058.
[21] J. Payne, C.M. Francis, K. Phillipps, A Field Guide to the Mammals of Borneo, The Sabah Society, Kota Kinabalu, 1985.
[22] J.F. Ducroz, V. Volobouv, L. Granjon, A molecular perspective systematic and evolution of the genus *Arvicanthis* (Rodentia: muridae): inferences from complete cytochrome *b* gene sequences, Mol. Phylogenetics Evol. 10 (1) (1998) 104–117.

[23] S. Kumar, G. Stecher, K. Tamura, MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets, Mol. Biol. Evol. 33 (2016) 1870–1874.
[24] J. Rozas, J.C. Sanchez-Delbarrio, X. Messeguer, R. Rozas, DnaSP: DNA polymorphism analyses by the coalescent and other methods, Bioinformation 19 (2003) 2496–2497.