Sains Malaysiana 50(4)(2021): 953-965
http://doi.org/10.17576/jsm-2021-5004-07

Mitochondrial Genome of *Rattus tiomanicus* (Rodentia: Muridae) and Molecular Phylogeny of Murinae

(Genom Mitokondria *Rattus tiomanicus* (Rodentia: Muridae) dan Filogeni Molekul Murinae)

SZE-LOOI SONG*, HOI-SEN YONG, PHAIK-EEM LIM & JI TAN

ABSTRACT

Rattus tiomanicus *is a murid rodent of considerable agricultural and public health importance in Southeast Asia. The whole mitochondrial genome of* R. tiomanicus *was sequenced by the Ion Torrent PGM platform. It had a total length of 16,309 bp, consisting of 13 protein-coding genes, two rRNA genes, 22 tRNA genes and two non-coding regions (L-strand replication origin and control region). Only TAA and incomplete T-stop codons were represented in the protein-coding genes. Of the tRNAs, tryptophan (W) had ACU anticodon. The cloverleaf structure for serine S1 (AGN) tRNA lacked the entire D-arm, while in lysine (K) tRNA, the DHU arm lacked the D-loop. Molecular phylogeny based on 15 mt-genes indicated* R. tiomanicus *having closest genetic affinity to* R. rattus *complex* (R. rattus, R. tanezumi)*. There were two major clades for the Murinae subfamily namely the Rattini tribe and the Apodemini, Murini and Hydromyini tribes. The whole mitogenome of* R. tiomanicus *will serve as a useful dataset for studying the systematics and phylogenetic relationships of the murid rodents.*

Keywords: Genomics; Murinae; phylogenetics

ABSTRAK

Rattus tiomanicus *ialah tikus murid yang penting dalam bidang kesihatan dan pertanian di Asia Tenggara. Seluruh genom mitokondria* R. tiomanicus *telah dijujuk oleh platfom Ion Torrent PGM. Ia mempunyai jumlah panjang 16,309 bp, terdiri daripada 13 gen yang mengekod protein, dua gen rRNA, 22 gen tRNA dan dua kawasan bukan pengekodan (origin replikasi bebenang L dan kawasan pengawalan). Hanya TAA dan kodon penamat tidak lengkap T berada dalam gen yang mengekod protein. Bagi gen tRNA, triptofan (W) mempunyai antikodon ACU. Struktur kelawar bagi tRNA serina S1 (AGN) kekurangan keseluruhan lengan-D, manakala lengan DHU dalam tRNA lisina (K) kekurangan gelung-D. Filogeni molekul berdasarkan gen-mt 15 menunjukkan* R. tiomanicus *mempunyai pertalian genetik yang paling dekat dengan* R. rattus *kompleks* (R. rattus, R. tanezumi)*. Dua klad utama untuk subfamili Murinae adalah untuk suku Rattini dan suku Apodemini, Murini serta Hydromyini. Keseluruhan mitogenom* R. tiomanicus *adalah bermanfaat untuk memberikan set data dalam mengkaji hubungan sistematik dan filogenetik bagi tikus murid.*

Kata kunci: Filogenetik; genom; Murinae

INTRODUCTION

Rattus tiomanicus is a murid rodent of considerable economic and public health importance, as a pest and carrier of disease pathogens, in Southeast Asia. It is a definitive host of the nematode parasites *Angiostrongylus cantonensis* (Yong & Eamsobhana 2013), *A. malaysiensis* (Eamsobhana 2014) and *Breinlia tinjili* (Purnomo & Bangs 1996) among others.

The Malaysian field rat *R. tiomanicus* is found in Thailand, Malaysia, Brunei, Indonesia and the Philippines (Musser & Carleton 2005), excluding the contentious report of transundaic distribution (Balakirev & Rozhnov 2012; Robins et al. 2014). The type locality is Tioman Island, off the east coast of Peninsular Malaysia. It is represented by many subspecies, both in the mainland as well as the offshore islands. It is a serious pest of plantation crops in Southeast Asia (Buckle et al. 1997).

The karyotype of *R. tiomanicus* from Tioman Island is identical to the mainland *R. t. jalorensis* (Yong et al. 1972). It is distinct from the closely related

congeneric taxa *R. rattus* and *R. argentiventer* (Yong 1969). These congeners are also readily distinguished by their immunoelectrophoretic serum-protein patterns (Yong 1968; Yong & Dhaliwal 1980). A study based on cytochrome *b* (*cob*) gene indicated that *R. tiomanicus* is closely related to *R. baluensis*, a montane endemic of northwestern Borneo (Aplin et al. 2011).

In addition to DNA nucleotide sequences (Latinne et al. 2013; Lecompte et al. 2008; Pagès et al. 2010; Robins et al. 2014, 2007), whole mitochondrial genomes have been used to infer phylogenetic relationships of the *Rattus* genus (Robins et al. 2010, 2008). Usage of single or several universal markers for phylogenetic analysis is becoming unfeasible as complete genome sequences will be able to generate a robust phylogeny. To date, there are 13 mitochondrial whole genomes (including the present *R. tiomanicus*) belonging to the *Rattus* genus deposited in GenBank. We report here the whole mitogenome of *R. tiomanicus* based on next-generation sequencing (NGS) and discuss the molecular phylogeny of Murinae.

Material and Methods

ETHICS STATEMENT

This research was approved by the UTAR Scientific and Ethical Review Committee (SERC) with approval number of U/SERC/70/2015 and was carried out as in guidelines stated by SERC.

SPECIMEN

The rat was collected by wire cage trap. It was sacrificed by anesthesia. Small pieces of liver were preserved in absolute ethanol and kept in freezer until use. The rat was identified based on the morphological descriptions from Medway (1983) and Yong et al. (1972).

MITOCHONDRIA ISOLATION AND DNA EXTRACTION

A small piece of the alcohol-preserved liver tissue was pressed onto a C-fold paper towel to remove excess ethanol before homogenisation. The mitochondria were isolated by standard differential centrifugation method and the mitochondrial DNA (mtDNA) was extracted using Mitochondrial DNA Isolation Kit (Abnova, Taipei, Taiwan) following the manufacturer's instructions. Final DNA was eluted using EB buffer instead of TE buffer to avoid interference of EDTA with the enzyme.

SAMPLE AND LIBRARY PREPARATION

The purified mtDNA was quantified with a Qubit 2.0 Fluorometer (Life Technologies, USA) using the Quant-iT dsDNA HS Assay Kit (Life Technologies, USA). A total of 100 ng of mtDNA was enzymatically fragmented using Ion Shear™ Plus Reagents (Life Technologies, USA) following the manufacturer's protocol. For 400-base-read library, incubation time was set to 6 min to yield median fragment size of 350 to 450 bp. Ion adapters were ligated to the fragmented mtDNA using the Ion Plus Fragment Library Kit (Life Technologies, USA) followed by nick repair. The library was size-selected using the 2.0% E-Gel SizeSelect™ Agarose Gel (Life Technologies, USA) following the manufacturer's recommendations. The quality, size distribution and concentration of the library were determined using the Agilent 2200 TapeStation (Agilent Technologies, Inc.) prior to template preparation to ensure library was within the recommended size range of the template kit used, as well as to minimise the polyclonal values of the sequencing results.

TEMPLATE PREPARATION

The library was diluted to a final concentration of 26 pM and subjected to emulsion PCR using the Ion PGM™ Template OT2 400 Kit and Ion OneTouch™ 2 System (Life Technologies, USA) to generate template-positive Ion Sphere™ Particles (ISPs) containing clonally amplified DNA. The template-positive ISPs were recovered and quantified using the Ion Sphere quality control kit (Life Technologies) and Qubit 2.0 fluorometer (Life Technologies) according to the manufacturer's recommendations. The percentage of template-positive ISPs fell within the optimal range of 10-30% was subsequently enriched with Dynabeads Myone streptavidin C1 beads (Life Technologies) using the Ion One-Touch™ ES (Life Technologies, USA) following the manufacturer's protocol.

PGM SEQUENCING

Chlorite cleaning procedures followed by a wash with 18MΩ water were performed before initialising the PGM Sequencer. The template-positive ISPs were loaded onto Ion 316™ V2 Chip and sequenced using the Ion PGM™ 400 Sequencing Kit (Life Technologies) on Ion Torrent Personal Genome Machine (PGM) System (Life Technologies) following the manufacturer's protocols.

SEQUENCE AND GENOME ANALYSIS

Raw sequences were extracted from the Torrent Server in FASTQ format and the quality of sequences was evaluated using the FastQC software (Andrews 2010). All the ambiguous nucleotides and reads with an average quality value (lower than Q20) were excluded from

further analysis. *De novo* assembly was performed using the CLC Genomic Workbench v.7.0.4 (https://www.qiagenbioinformatics.com/) and contigs greater than 15 kbp were subjected to BLAST (Altschul et al. 1990) alignment against the nucleotide database at National Center for Biotechnology Information (NCBI).

The three domains (ETAS, CD, and CSB) and main subsequences of the control region (CR, D-loop) were ascertained by reference to those reported for *R. norvegicus* and other rodents (Abhyankar et al. 2009; Gemmell et al. 1996; Larizza et al. 2002; Sbisà et al. 1997; Silva et al. 2011).

MITOGENOME IDENTIFICATION, ANNOTATION AND VISUALISATION

The assembled contig identified as mitogenome was manually examined for repeats at the beginning and end of the sequence to infer circularity. The mitogenome was then annotated with MITOS (http://mitos.bioinf.uni-leipzig.de/index.py) (Bernt et al. 2012) followed by manual validation of the coding regions using the NCBI ORF Finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html). The sequin file generated from MITOS was edited and submitted to NCBI according to ORF Finder result (NCBI GenBank accession number KP876560 and NC_029888). The circular mitogenome of *R. tiomanicus* was visualised with Blast Ring Image Generator (BRIG) (Alikhan et al. 2011).

PHYLOGENETIC ANALYSIS

Mitogenome sequences (36 species) of Muridae family available in GenBank were used to construct phylogenetic trees. *Pteromys volans* NC_019612 (Ryu et al. 2013) and *Ratufa bicolor* NC_023780 (Kong et al. 2015) of the Sciuridae family were included as outgroup taxa. The 15 mt-gene sequences were aligned using MAFFT multiple sequence alignment software v.7 (Katoh & Standley 2013) and subsequently edited and trimmed using BioEdit v.7.0.5.3 (Hall 1999). Kakusan v.3 (Tanabe 2007) was used to determine the best-fit nucleotide substitution models for maximum likelihood (ML) and Bayesian (BI) analyses selected using the corrected Akaike Information Criterion (Akaike 1973) and the Bayesian Information Criterion (Schwarz 1978), respectively. Phylograms of 15 mt-genes were constructed using TreeFinder (Jobb et al. 2004) prior to the annotations of bootstrap values (BP) generated *via* 1,000 ML bootstrap replicates. Bayesian analyses on the concatenated amino acid sequence data for all 15 mt-genes were conducted using the Markov chain Monte Carlo (MCMC) method *via* Mr. Bayes v.3.1.2 (Huelsenbeck & Ronquist 2001). Two independent runs of 2×10^6 generations with four chains were performed, with trees sampled every 200th generation. Likelihood values for all post-analysis trees and parameters were evaluated for convergence and burn-in using the 'sump' command in MrBayes and the computer program Tracer v.1.5 (http://tree.bio.ed.ac.uk/software/tracer/). The first 200 trees from each run were discarded as burn-in (where the likelihood values were stabilized prior to the burn-in), and the remaining trees were used for the construction of a 50% majority-rule consensus tree. FigTree v.1.4 was used to view and edit the final phylogenetic trees (Rambaut 2012).

RESULTS AND DISCUSSION

GENOME FEATURES

The *Rattus* genus is represented by some 64 species in the world (Musser & Carleton 2005). To date, the complete mitogenomes of 13 species (including the present *R. tiomanicus*) have been sequenced and available in GenBank. This made up nearly one-third of the complete mitogenomes of 36 species for the Muridae family. Most of these complete mitogenomes have been sequenced by means of the long polymerase chain reaction (long-PCR) technique (Robins et al. 2010, 2008; Wang et al. 2015; Yong et al. 2016; Zhu et al. 2016). In this study, we sequenced and reported for the first time the complete mitogenome of *R. tiomanicus* by NGS using the Ion Torrent PGM platform. The results were consistent with reported rodent mitogenomes based on long-PCR. In comparison to other methods such as long-PCR, Sanger sequencing and Illumina platform, Ion Torrent PGM platform is performed by using semiconductor chip which is fast, simple and cost effective. Similar to other NGS platforms, Ion Torrent PGM produces large amounts of DNA sequence data in a short period of time.

Next-generation sequencing on Ion Torrent PGM platform generated a total of 1,440,186 sequence reads, with an average length of 159 bases and a total of 229,400,258 bases. Removal of low quality sequence, ambiguous nucleotides and sequences shorter than 50 nucleotides resulted in 13,709 contigs with 5,194,894 bases. The contig maximum length was 16,309 bp and N50 was 383 bp. The total GC content was 46.7%, with base composition of 34.2% A, 27.8% T, 12.5% G, and 25.5% C. The mitogenome of *R. tiomanicus* was 16,309 bp long consisting of 37 genes (13 PCGs, two rRNA genes and 22 tRNA genes) as well as two non-coding regions (L-strand replication origin and control region) (Table 1 & Figure 1). Spacing sequences ranged from 1 to 5 bp in 14 regions. The overlaps in seven regions ranged from 1 to 43 bp, the largest being between *atp8* and *atp6* (Table 1). Nine genes (one protein-coding gene *nad6*; eight tRNA genes) were

located on the light-strand. The L-strand replication origin (O_L) was located in the WANCY cluster between *trnN* and *trnC* (Table 1 & Figure 1), with 31 bp as in other species of the *Rattus* genus (Table 2). It had two base substitutions - a transversion (T to A) at position 14 (also in *R. rattus* and *R. tanezumi*), and a transition (C to T) at position 15 compared to other congeners (Table 2).

TABLE 1. Characteristics of the mitochondrial genome of *Rattus tiomanicus*. Three nucleotides in parentheses denotes the anticodon

Gene	Location	Strand	Size (bp)	Intergenic sequence	Start/Stop codon
trnF(gaa)	1 – 69	H	69		
rrnS	70 – 1031	H	962		
trnV(tac)	1032 – 1099	H	68		
rrnL	1100 – 2665	H	1566		
trnL2(taa)	2666 – 2740	H	75		
nad1	2741 – 3695	H	955		GTG/T
trnI(gat)	3696 – 3764	H	69	-3	
trnQ(ttg)	3762 – 3832	L	71	3	
trnM(cat)	3836 – 3904	H	69		
nad2	3905 – 4940	H	1036		ATA/T
trnW(tca)	4941 – 5006	H	66	1	
trnA(tgc)	5008 – 5076	L	69	2	
trnN(gtt)	5079 – 5150	L	72		
O_L	5151 – 5181	L	31		
trnC(gca)	5182 – 5248	L	67	3	
trnY(gta)	5252 – 5317	L	66	1	
cox1	5319 – 6863	H	1545	-3	ATG/TAA
trnS2(tga)	6861 – 6929	L	69	3	
trnD(gtc)	6933 – 7000	H	68	3	
cox2	7002 – 7685	H	684	3	ATG/TAA
trnK(ttt)	7689 – 7752	H	64	1	
atp8	7754 – 7957	H	204	-43	ATG/TAA
atp6	7915 – 8595	H	681	-1	ATG/TAA
cox3	8595 – 9378	H	784		ATG/T
trnG(tcc)	9379 – 9446	H	68		
nad3	9447 – 9794	H	348	1	ATC/TAA
trnR(tcg)	9796 – 9863	H	68	2	
nad4l	9866 – 10162	H	297	-7	ATG/TAA
nad4	10156 – 11533	H	1378		ATG/T
trnH(gtg)	11534 – 11601	H	68		
trnS1(gct)	11602 – 11660	H	59	-1	
trnL1(tag)	11660 – 11730	H	71		
nad5	11731 – 13560	H	1830	-23	ATA/TAA
nad6	13538 – 14056	L	519		ATG/TAA
trnE(ttc)	14057 – 14125	L	69	5	
cob	14131 – 15273	H	1143	1	ATG/TAA
trnT(tgt)	15275 – 15341	H	67	2	
trnP(tgg)	15344 – 15411	L	68		
Control region	15412 – 16309	H	898		

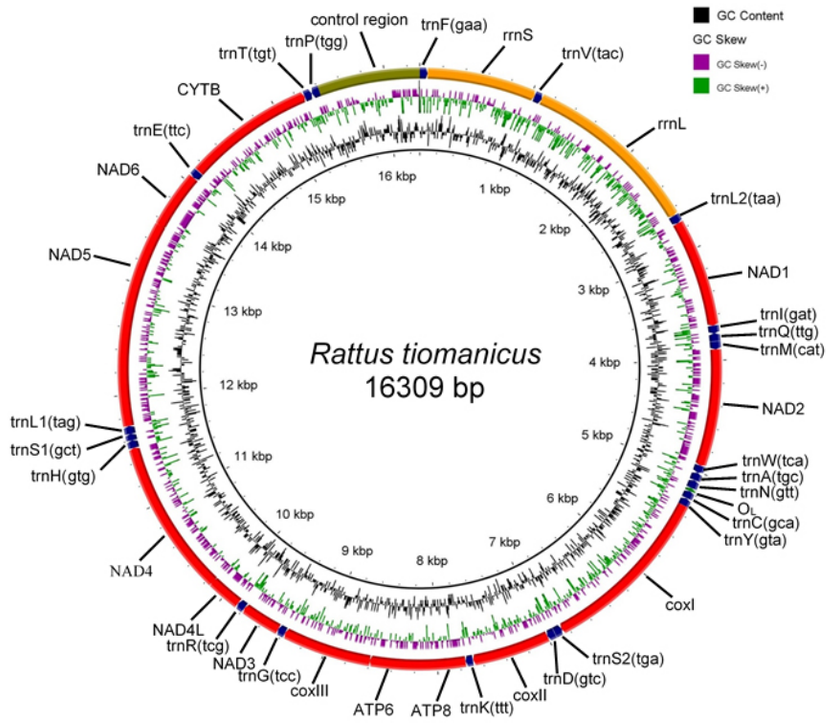

FIGURE 1. Complete mitogenome of *Rattus tiomanicus* with BRIG visualisation. It shows the protein coding genes, rRNAs, tRNAs and non-coding regions. GC skew is shown on the outer surface of the ring whereas GC content is shown on the inner surface

TABLE 2. Alignment of the L-strand Replication Origin sequences of *R. tiomanicus* and congeners with reference to *R. norvegicus*. Base substitutions at positions 14 and 15 with reference to *R. norvegicus* are highlighted

Species	Nucleotide sequence
R. tiomanicus	CTTCTCCCGCCTAATAGAAAAGAGGCGGGAG
R. rattus NC_012374	CTTCTCCCGCCTAACAGAAAAGAGGCGGGAG
R. tanezumi NC_011638	CTTCTCCCGCCTAACAGAAAGGGGGCGGGAG
R. norvegicus KF011917	CTTCTCCCGCCTATCAGAAAAGAGGCGGGAG
R. exulans NC_012389	CTTCTCCCGCCTATCAGAAAAGGGGCGGGAG
R. fuscipes NC_014867	CTTCTCCCGCCTATCAGAAAAGAGGCGGGAG
R. leucopus NC_014855	CTTCTCCCGCCTATCAGAAAAGAGGCGGGAG
R. lutreolus NC_014858	CTTCTCCCGCCTATCAGAAAAGAGGCGGGAG
R. niobe NC_023347	CTTCTCCCGCCTATCAGAAAAGAGGCGGGAG
R. praetor NC_012461	CTTCTCCCGCCTATCAGAAAAGAGGCGGGAG
R. sordidus NC_014871	CTTCTCCCGCCTATCAGAAAAGAGGCGGGAG
R. tunneyi NC_014861	CTTCTCCCGCCTATCAGAAAAGAGGCGGGAG
R. villosissimus NC_014864	CTTCTCCCGCCTATCAGAAAAGAGGCGGGAG

The control region (898 bp) was flanked by *trnP* and *trnF* (Figure 1). It conformed to the general scheme of the D-loop structure in rodents, comprising the ETAS (extended termination-associated sequences) domain (ETAS1 and ETAS2), CD - central domain (subsequences A, B and C), and CSB (conserved sequence block) domain (CSB1, CSB2 and CSB3) (Figure 2).

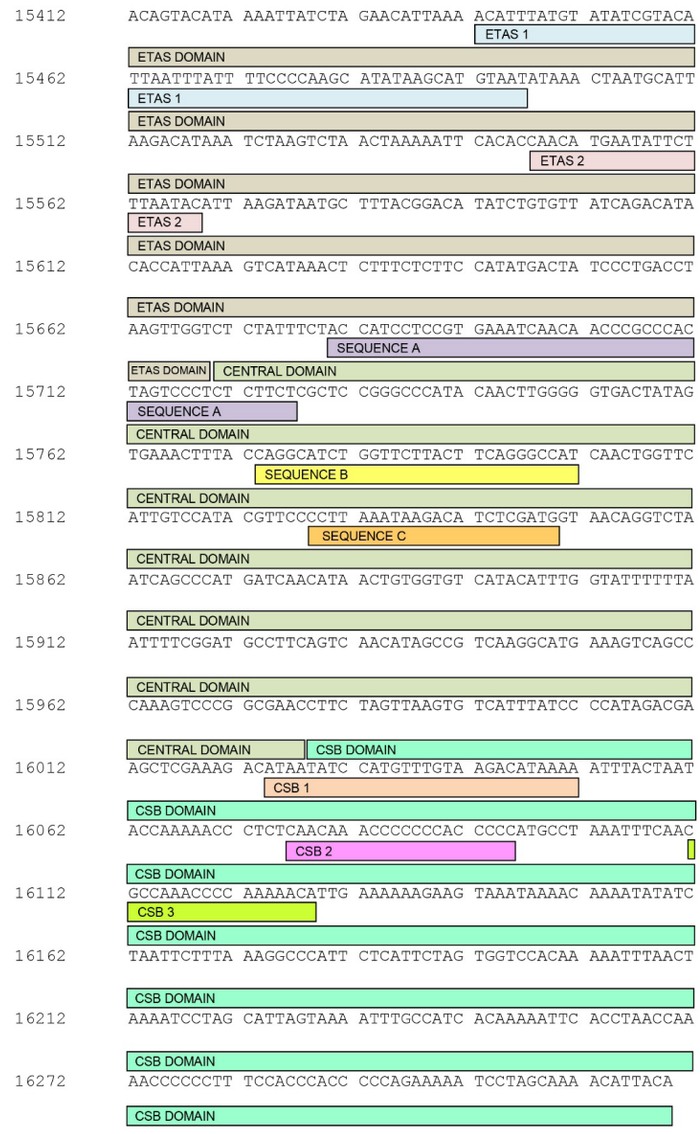

FIGURE 2. Control region in the mitogenome of *Rattus tiomanicus*. ETAS (bp position: 15412-15669), extended termination-associated sequences; Central domain (bp position: 15670-15977); CSB (bp position: 15978-16309), conserved sequence block. ETAS1: 15442-15496; ETAS2: 15547-15568; Sequence A: 15680-15726; Sequence B: 15773-15801; Sequence C: 15828-15849; CSB 1: 16024-16051; CSB 2: 16076-16095; CSB 3: 16111-16228

The commonest start codon was ATG (in nine protein-coding genes - *cox1*, *cox2*, *atp8*, *atp6*, *cox3*, *nad4l*, *nad4*, *nad6*, *cob*), followed by two for ATA (*nad2*, *nad5*), and one each for ATC (*nad3*) and GTG (*nad1*). Nine protein-coding genes had TAA stop codon while the remaining four genes (*nad1*, *nad2*, *cox3*, *nad4*) had incomplete T– stop codon (Table 1).

Of the tRNAs, *trnW* had ACU anticodon instead of the canonical Trp tRNA ACC anticodon (Table 1 & Figure 3). The cloverleaf structure for *trnS1* lacked the entire D-arm, while in *trnK* the DHU arm lacked the D-loop (Figure 3). The number of base pairs in the DHU-stem ranged from 2 to 4 except none in *trnS1* (Figure 3 & Table 3). *trnS2* had 2 bp for the D-stem. All the TΨC-stems had 5 base pairs except 4 bp in *trnW*.

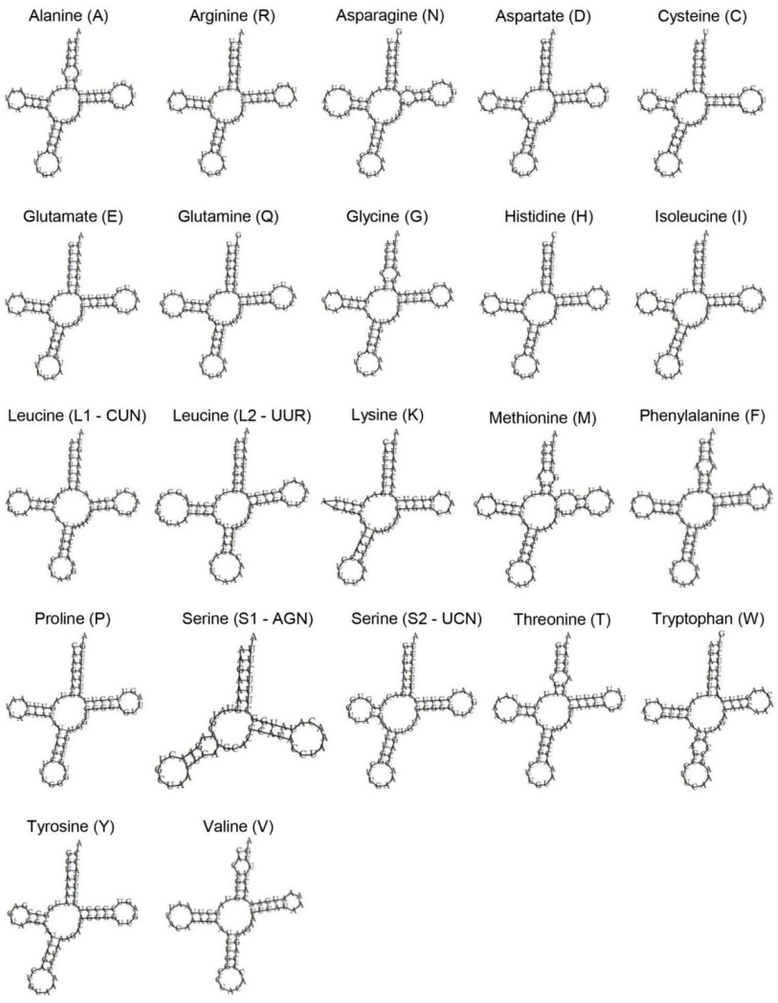

FIGURE 3. Cloverleaf structure of the 22 inferred tRNAs in the mitogenome of *Rattus tiomanicus*. Note atypical secondary structure of lysine with the DHU arm lacking the D-loop and serine S1 without the entire D-arm

TABLE 3. Number of base pairs in DHU-stem and TΨC-stem of mt-tRNAs of *Rattus tiomanicus*

tRNA	DHU-stem	TΨC-stem
Alanine	4	5
Arginine	4	5
Asparagine	3	5#
Aspartate	4	5
Cysteine	3	5
Glutamate	3	5
Glutamine	3	5
Glycine	4	5
Histidine	4	5
Isoleucine	3	5
Leucine L1 (CUN)	4	4
Leucine L2 (UUR)	4	5
Lysine	4*	5
Methionine	4	5#
Phenylalanine	4	5
Proline	4	5
Serine S1 (AGN)	absent	5
Serine S2 (UCN)	2	5
Threonine	3	5
Tryptophan	4	4
Tyrosine	3	5
Valine	4	5

*without D-loop; #with small inner loop

The size of the *R. tiomanicus* mitogenome (16,309 bp) is similar to other species of the *Rattus* genus (range of 16,292 to 16,310 bp) and other murid genera (range of 16,217 to 16,351 bp) that are available in GenBank. In this *R. tiomanicus* mitogenome, only TAA and incomplete T-stop codons were represented in the protein-coding genes (Table 1 & Figure 2). The stop codons in *R. tiomanicus* differ from other rodent mitogenomes which possess additionally TAG stop codon (Chao et al. 2014; Jiang et al. 2012; Kim & Lee 2016; Kong et al. 2015) and other stop codons - AGG and AGA (Ryu et al. 2013); and TAT, CAT and CTT (Zhao et al. 2014). The incomplete T–stop codons can be converted to TAA by post-translational polyadenylation (Ojala et al. 1981).

As in other vertebrates, the *R. tiomanicus* mitogenome has three clusters of characteristic tRNAs (Figure 1) which were IQM (isoleucine, glutamate, and methionine), WANCY (tryptophan, alanine, aspartic acid, cysteine, and tyrosine) and HSL (histidine, serine, and leucine) (Pereira 2000). Among the unusual features of the tRNAs are the presence of ACU anticodon in *trnW*, absence of the entire D-arm for *trnS1*, and absence of D-loop in the DHU arm of *trnK*. The ACU anticodon is present in 12 other congeners of the *Rattus* genus (Figure

4). The atypical cloverleaf structures of *trnS1* and *trnK* are also found in the Arvicolinae rodent *Microtus fortis calamorum* (Jiang et al. 2012). Indeed, the absence of DHU arm in *trnS1* is common in all vertebrates and a reduced D-arm in *trnK* is shared by placental and marsupial mammals (Pereira 2000).

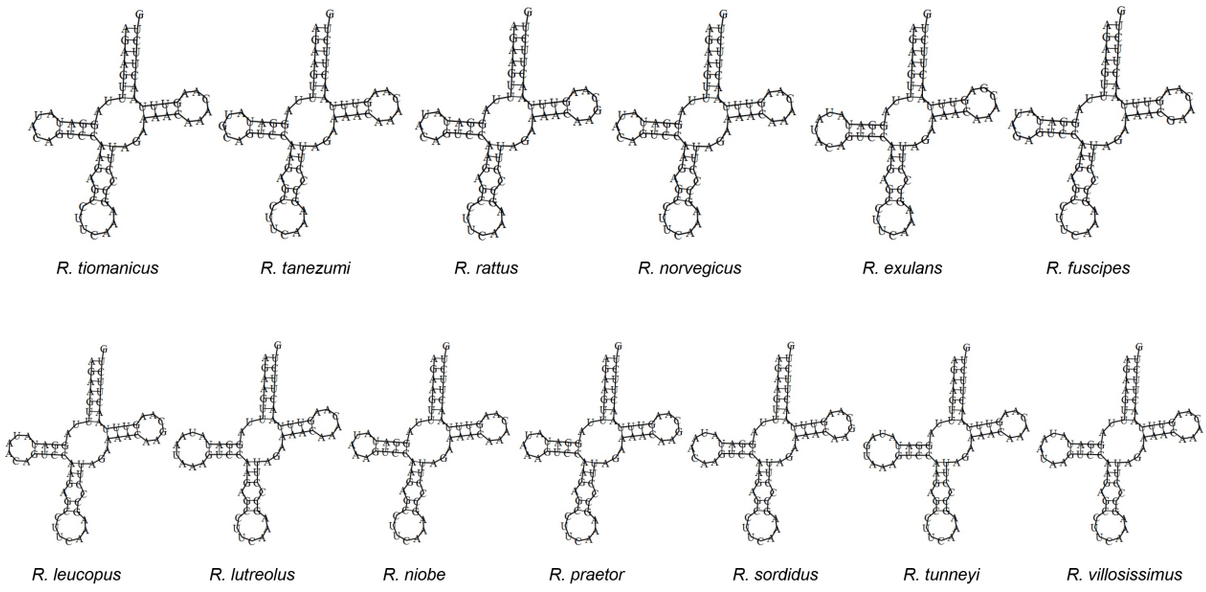

FIGURE 4. Secondary structure of tryptophan sequences of *Rattus tiomanicus* and congeners. It shows identical ACU anticodon (opal suppressor) instead of the canonical Trp tRNA ACC anticodon

PHYLOGENETIC RELATIONSHIPS WITHIN MURINAE

The total length of the aligned sequences of 15 mt-genes (13 PCGs, two rRNA genes) was 14,222 bp and the selected models used for maximum likelihood (ML) and Bayesian Inference (BI) analyses were GTR+Gamma and SYM+Gamma, respectively. The molecular phylogeny of *R. tiomanicus* in relation to other taxa of the Murinae subfamily is depicted in Figure 5. Most of the nodes were well-supported. There were two major clades for the Murinae subfamily namely the Rattini tribe comprising *Rattus* and *Dacnomys* divisions, and Apodemini tribe (*Apodemus* division), Murini tribe (*Mus* division) and Hydromyini tribe (*Psuedomys* division), with *Apodemus* division and *Mus* division having closer affinity than *Pseudomys* division. *R. tiomanicus* showed closer genetic affinity to *R. rattus* and *R. tanezumi* than to *R. exulans* and *R. norvegicus*.

Phylogenetic analysis based on DNA nucleotide sequences, particularly D-loop, cytochrome *b* (*cob*) and cytochrome oxidase I (*cox1*), assigned *R. tiomanicus* to a phylogenetic group distinct from *Rattus rattus* species complex (Latinne et al. 2013; Pagès et al. 2010; Robins et al. 2014, 2007). It is closely related to *R. baluensis* forming Lineage VI in a clade with Lineage IV (*Rattus rattus*) and Lineage V (*Rattus sakeratensis*) (Aplin et al. 2011).

In the present study, the molecular phylogeny based on 15 mt-genes (Figure 5) showed two lineages in the *Rattus* division which were Asian and Island Southeast Asian *Rattus* species, and Australo-Papuan *Rattus* species. This phylogenetic relationship is in accordance with the findings based on D-loop, *cob* and *cox1* sequences (Robins et al. 2007), D-loop and *cox1* sequences (Robins et al. 2014) and whole mitochondrial genome (Robins et al. 2010). The closer affinity of *R. tiomanicus* with *R. rattus* and *R. tanezumi* than to *R. exulans* and *R. norvegicus* is congruent with findings based on D-loop, *cob* and *cox1* sequences (Latinne et al. 2013; Pagès et al. 2010; Robins et al. 2014, 2007).

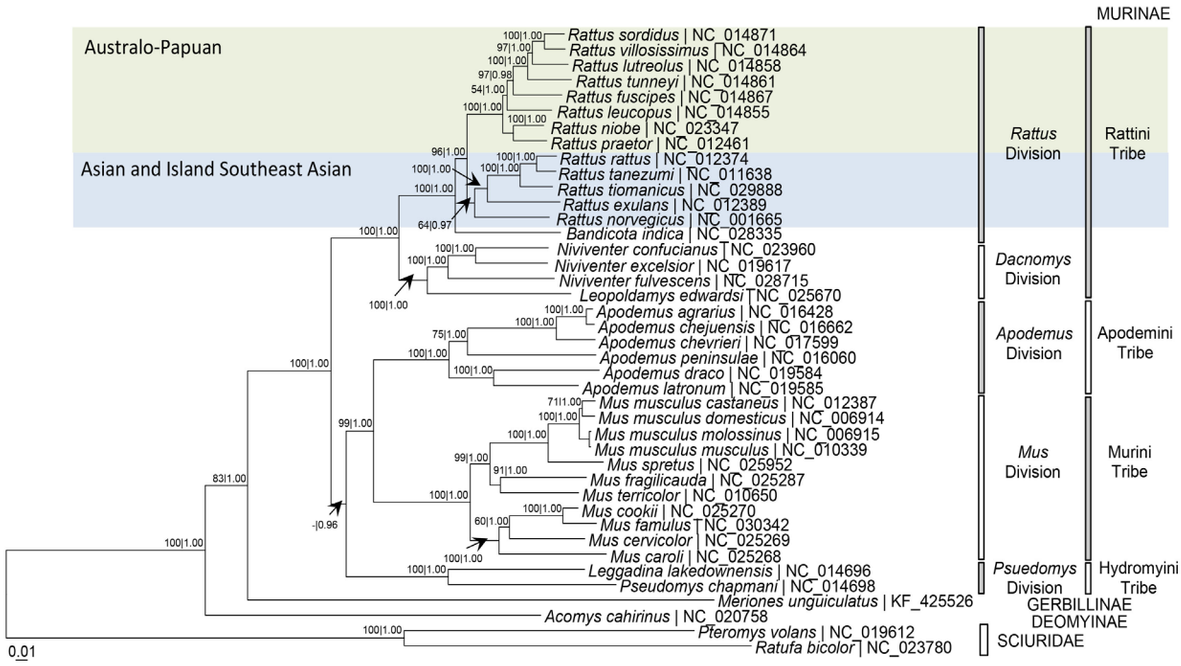

FIGURE 5. BI and ML tree based on 15 mt-genes (13 PCGs, two rRNA genes) of the whole mitogenomes of murid rodents. Numeric values at the nodes are Bayesian posterior probabilities/ML bootstrap

The phylogenetic relationships of the Australo-Papuan *Rattus* species based on 15 mt-genes (Figure 5) were in general agreement with the findings based on D-loop, *cob* and *cox1* sequences (Robins et al. 2007) and D-loop and *cox1* sequences (Robins et al. 2014), but not congruent with the findings based on *cob*, GHR (growth hormone receptor exon 10) and IRPB (interphotoreceptor retinoid binding protein exon1) sequences (Fabre et al. 2013). Our results were also in general agreement with the findings based on whole mitogenome datasets of seven species (Robins et al. 2010). A major difference is the inclusion of *R. niobe* mitogenome in our analysis resulting in the close affinity of *R. praetor* and *R. niobe* instead of *R. praetor* with *R. fuscipes*. However, the topology of the phylogenetic tree may change when mitogenomes of other taxa become available for study. To date, the study based on D-loop and *cox1* sequences indicated that *R. praetor*, *R. niobe,* and *R. verecundus* each likely encompass more than one species (Robins et al. 2014).

Our present findings of *Bandicota indica* being basal to the *Rattus* genus in the *Rattus* Division and of *Leopoldamys edwardsi* being basal to the *Niviventer* genus in the *Dacnomys* Division (Figure 5) concurred with earlier studies (Wang et al. 2015; Yong et al. 2016).

The Murinae phylogeny in the present study was congruent with that based on *cob*, IRBP and GHR genes (Lecompte et al. 2008) although the datasets are different. The phylogeny of the *Apodemus* genus based on 15 mt-genes (Figure 5) of available whole mitogenomes shows good concordance with that based on *cob* sequences (Liu et al. 2012), comprising two subgroups in the *Apodemus* group namely *agrarius* subgroup consisting of *A. agrarius*, *A. chejuensis*, *A. chevrieri,* and *A. peninsulae*, and *draco* subgroup consisting of *A. draco* and *A. latronum*. Our study showed additionally close affinity of *A. agrarius* and *A. chejuensis* where the taxon *chejuensis* was earlier treated as a subspecies of *A. agrarius* but found to be valid species based on differences in mitochondrial control region sequences and morphology (Koh et al. 2000). The topology of the phylogeny of the *Apodemus* group based on 15 mt-genes (Figure 5) and *cob* sequences (Liu et al. 2012) differed from that inferred from *cob*, IRBP, RAGI (recombination activating gene 1), I7 (an olfactory receptor) and vWF (von Willebrand factor) genes (Suzuki et al. 2008) which showed closer affinity of *A. peninsulae* to the *draco* subgroup.

In the present study, 11 taxa of the *Mus* genus with complete mitogenomes have been included for analysis (Figure 5). The overall phylogenetic relationship was congruent with that inferred from *cob*, 12S, B2m (B2-microglobulin), Zp3 (zona pellucida-3), Tcp1 (t-complex polypeptide-1), Sry (sex determining locus), Smcx

(member of jumonji family of transcription factors) and Smcy (selected mouse cDNA Y) sequences (Tucker et al. 2005). Our analysis included additionally *M. famulus*, *M. fragilicauda* and *M. terricolor*. *M. fragilicauda* and *M. terricolor* are sister taxa and form a clade with *M. musculus* and *M. spretus*, while *M. famulus* is sister to *M. cookii* in the lineage comprising also *M. caroli* and *M. cervicolor* (Figure 5).

At the subfamily level based on 15 mt-genes (Figure 5), Gerbillinae showed closer affinity to Murinae than Deomyinae. This differed from the findings of Gerbillinae and Deomyinae (Acomyinae) being sister taxa inferred from slower evolving IRBP gene sequences (Jansa & Weksler 2004) and nuclear protein-coding LCAT (lecithin cholesterol acyl transferase) and vWF genes (Michaux et al. 2001). More extensive taxa sampling of some under-represented and unrepresented subfamilies was needed to resolve the phylogenetic relationships of the murid subfamilies.

Conclusion

In summary, we have successfully sequenced the whole mitochondrial genome of *R. tiomanicus* by next-generation sequencing. The genome features were similar to other rodents except that only TAA and incomplete T–stop codons were represented in the protein-coding genes, and the presence of ACU anticodon in *trnW*. Based on 15 mt-genes of the mitogenome, *R. tiomanicus* showed closest genetic affinity to *R. rattus* and *R. tanezumi*. There were two major clades for the Murinae subfamily namely Rattini tribe, and Apodemini, Murini and Hydromyini tribes. The tribes/divisions of the Murinae subfamily were monophyletic. At the subfamily level, Gerbillinae showed closer affinity to Murinae than to Deomyinae. The whole mitogenome of *R. tiomanicus* will serve as a useful dataset for studying the systematics and phylogenetic relationships of the many subspecies of *R. tiomanicus* in particular, and murid rodents in general.

Data Availability Statement

The whole mitochondrial genome sequence of *R. tiomanicus* is available in GenBank database (accession number: KP876560 and NC_029888). The authors have declared no conflicts of interest.

Acknowledgements

This study was funded in part by MOHE-HIR Grant (H-50001-00-A000025) and the University of Malaya (H-5620009). We thank our institutions for providing various research facilities and other support.

References

Abhyankar, A., Park, H.B., Tonolo, G. & Luthman, H. 2009. Comparative sequence analysis of the non-protein-coding mitochondrial DNA of inbred rat strains. *PLoS ONE* 4(12): e8148.

Akaike, H. 1973. Information theory and an extension of the maximum likelihood principle. In *Second International Symposiumon Information Theory*, edited by Csaki, F. & Petrov, B.N. Budapest: Akademia Kiado. pp. 267-281.

Alikhan, N.F., Petty, N.K., Ben Zakour, N.L. & Beatson, S.A. 2011. BLAST Ring Image Generator (BRIG): Simple prokaryote genome comparisons. *BMC Genomics* 12(1): 402.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. 1990. Basic local alignment search tool. *Journal of Molecular Biology* 215(3): 403-410.

Andrews, S. 2010. *FastQC: A Quality Control Tool for High Throughput Sequence Data*. Cambridge: Babraham Institute.

Aplin, K.P., Suzuki, H., Chinen, A.A., Chesser, R.T., Ten Have, J., Donnellan, S.C., Austin, J., Frost, A., Gonzalez, J.P., Herbreteau, V., Catzeflis, F., Soubrier, J., Fang, Y.P., Robins, J., Matisoo-Smith, E., Bastos, A.D., Maryanto, I., Sinaga, M.H., Denys, C., Van Den Bussche, R.A., Conroy, C., Rowe, K. & Cooper, A. 2011. Multiple geographic origins of commensalism and complex dispersal history of black rats. *PLoS ONE* 6(11): e26357.

Balakirev, A. & Rozhnov, V.V. 2012. Contribution to the species composition and taxonomic status of some *Rattus* inhabiting Southern Vietnam and Sundaland. *Russian Journal of Teriology* 11(1): 33-45.

Bernt, M., Donath, A., Juhling, F., Externbrink, F., Florentz, C., Fritzsch, G., Pütz, J., Middendorf, M. & Stadler, P.F. 2012. MITOS: Improved de novo metazoan mitochondrial genome annotation. *Molecular Phylogenetics and Evolution* 69(2): 313-319.

Buckle, A.P., Chia, T.H., Fenn, M.G.P. & Visvalingam, M. 1997. Ranging behaviour and habitat utilisation of the Malayan wood rat (*Rattus tiomanicus*) in an oil palm plantation in Johore, Malaysia. *Crop Protection* 16(5): 467-473.

Chao, Q.J., Li, Y.D., Geng, X.X., Zhang, L., Dai, X., Zhang, X., Li, J. and Zhang, H.J. 2014. Complete mitochondrial genome sequence of *Marmota himalayana* (Rodentia: Sciuridae) and phylogenetic analysis within Rodentia. *Genetics and Molecular Research* 13(2): 2739-2751.

Eamsobhana, P. 2014. *The Rat Lungworm Angiostrongylus Cantonensis: Parasitology, Genetics and Molecular Phylogeny* (2nd ed.). Bangkok: Aksorn Graphic and Design Publishing House.

Fabre, P.H., Pagès, M., Musser, G.G., Fitriana, Y.S., Fjeldså, J., Jennings, A., Jønsson, K.A., Kennedy, J., Michaux, J., Semiadi, G., Supriatna, N. & Helgen, K.M. 2013. A new genus of rodent from Wallacea (Rodentia: Muridae: Murinae: Rattini), and its implication for biogeography and Indo-Pacific Rattini systematics. *Zoological Journal of the Linnean Society* 169(2): 408-447.

Gemmell, N.J., Western, P.S., Watson, J.M. & Graves, J.A. 1996. Evolution of the mammalian mitochondrial control region - comparisons of control region sequences between monotreme and therian mammals. *Molecular Biology and Evolution* 13(6): 798-808.

Hall, T.A. 1999. BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* 41: 95-98.

Huelsenbeck, J.P. & Ronquist, F. 2001. MrBayes: Bayesian inference of phylogenetic trees. *Bioinformatics* 17(8): 754-755.

Jansa, S.A. & Weksler, M. 2004. Phylogeny of muroid rodents: Relationships within and among major lineages as determined by IRBP gene sequences. *Molecular Phylogenetics and Evolution* 31(1): 256-276.

Jiang, X., Gao, J., Ni, L., Hu, J., Li, K., Sun, F., Xie, J., Bo, X., Gao, C., Xiao, J. & Zhou, Y. 2012. The complete mitochondrial genome of *Microtus fortis calamorum* (Arvicolinae, Rodentia) and its phylogenetic analysis. *Gene* 498(2): 288-295.

Jobb, G., von Haeseler, A. & Strimmer, K. 2004. Treefinder: A powerful graphical analysis environment for molecular phylogenetics. *BMC Evolutionary Biology* 4(1): 18.

Katoh, K. & Standley, D.M. 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution* 30(4): 772-780.

Kim, E.B. & Lee, S.G. 2016. The complete mitochondrial genome of the Mongolian gerbil, *Meriones unguiculatus* (Rodentia: Muridae: Gerbillinae). *Mitochondrial DNA* 27(2): 1457-1458.

Koh, H.S., Lee, W.J. & Kocher, T.D. 2000. The genetic relationships of two subspecies of striped field mice, *Apodemus agrarius coreae* and *Apodemus agrarius chejuensis*. *Heredity* 85(1): 30-36.

Kong, L., Wang, W., Cong, H., Liu, Z. & Li, Y. 2015. Complete mitochondrial genome of the black giant squirrel *Ratufa bicolor* (Rodentia: Sciuridae). *Mitochondrial DNA* 26(5): 759-760.

Larizza, A., Pesole, G., Reyes, A., Sbisà, E. & Saccone, C. 2002. Lineage specificity of the evolutionary dynamics of the mtDNA D-loop region in rodents. *Journal of Molecular Evolution* 54(2): 145-155.

Latinne, A., Waengsothorn, S., Rojanadilok, P., Eiamampai, K., Sribuarod, K. &Michaux, J.R. 2013. Diversity and endemism of Murinae rodents in Thai limestone karsts. *Systematics and Biodiversity* 11(3): 323-344.

Lecompte, E., Aplin, K., Denys, C., Catzeflis, F., Chades, M. & Chevret, P. 2008. Phylogeny and biogeography of African Murinae based on mitochondrial and nuclear gene sequences, with a new tribal classification of the subfamily. *BMC Evolutionary Biology* 8(1): 199.

Liu, Q., Chen, P., He, K., Kilpatrick, C.W., Liu, S.Y., Yu, F.H. & Jiang, X.L. 2012. Phylogeographic study of *Apodemus ilex* (Rodentia: Muridae) in Southwest China. *PLoS ONE* 7(2): e31453 012.

Medway, L. 1983. *The Wild Mammals of Malaya (Peninsular Malaysia and Singapore)*. Kuala Lumpur: Oxford University Press.

Michaux, J., Reyes, A. & Catzefkis, F. 2001. Evolutionary history of the most speciose mammals: Molecular phylogeny of muroid rodents. *Molecular Biology and Evolution* 18(11): 2017-2031.

Musser, G.G. & Carleton, M.D. 2005. Superfamily Muroidea. In *Mammal Species of the World a Taxonomic and Geographic Reference*, edited by Wilson, D.E. & Reeder, D.M. Baltimore: Johns Hopkins University Press. pp. 894-1531.

Ojala, D., Montoya, J. & Attardi, G. 1981. tRNA punctuation model of RNA processing in human mitochondria. *Nature* 290: 470-474.

Pagès, M., Chaval, Y., Hernerteau, V., Waengsothorn, S., Cosson, J.F., Hugot, J.P., Morand, S. & Michaux, J. 2010. Revisiting the taxonomy of the Rattini tribe: A phylogeny based determination of species boundaries. *BMC Evolutionary Biology* 10(1): 184.

Pereira, S.L. 2000. Mitochondrial genome organization and vertebrate phylogenetics. *Genetics and Molecular Biology* 23(4): 745-752.

Purnomo & Bangs, M.J. 1996. *Breinlia tinjili* sp. n. (Filarioidea: Onchocercidae), from the Malaysian Field Rat, *Rattus tiomanicus,* on Tinjil Island, West Java, Indonesia. *Journal of the Helminthological Society of Washington* 63: 93-97.

Rambaut, A. 2012. *FigTree*. Edinburgh: University of Edinburgh.

Robins, J.H., Tintinger, V., Aplin, K.P., Hingston, M., Matisoo-Smith, E., Penny, D. & Lavery, S.D. 2014. Phylogenetic species identification in *Rattus* highlights rapid radiation and morphological similarity of New Guinean species. *PLoS ONE* 9(5): e98002.

Robins, J.H., McLenachan, P.A., Phillips, M.J., McComish, B.J., Matisoo-Smith, E. & Rosset, H.A. 2010. Evolutionary relationships and divergence times among the native rats of Australia. *BMC Evolutionary Biology* 10(1): 375.

Robins, J.H., McLenachan, P.A., Phillips, M.J., Craig, L., Ross, H.A. & Matisoo-Smith, E. 2008. Dating of divergences within the *Rattus* genus phylogeny using whole mitochondrial genomes. *Molecular Phylogenetics and Evolution* 49(2): 460-466.

Robins, J.H., Hingston, M., Matisoo-Smith, E. & Ross, H.A. 2007. Identifying *Rattus* species using mitochondrial DNA. *Molecular Ecology Notes* 7(5): 717-729.

Ryu, S.H., Kwak, M.J. & Hwang, U.W. 2013. Complete mitochondrial genome of the Eurasian flying squirrel *Pteromys volans* (Sciuromorpha, Sciuridae) and revision of rodent phylogeny. *Molecular Biology Reports* 40(2): 1917-1926.

Sbisà, E., Tanzariello, F., Reyes, A., Pesole, G. & Saccone, C. 1997. Mammalian mitochondrial D-loop region structural analysis: identification of new conserved sequences and their functional and evolutionary implications. *Gene* 205(1-2): 125-140.

Schwarz, G. 1978. Estimating the dimension of a model. *Annals of Statistics* 6(2): 461-464.

Silva, A., Artiles, A., Suarez, W. & Silva, G. 2011. Sequence and structure of the mitochondrial control region of the Cuban rodent *Capromys pilorides* (Rodentia: Capronyidae). *Biotecnología Aplicada* 28(3): 136-141.

Suzuki, H., Filippucci, M.G., Chelomina, G.N., Sato, J.J., Serizawa, K. & Nevo, E. 2008. A biogeographic view of *Apodemus* in Asia and Europe inferred from nuclear and mitochondrial gene sequences. *Biochemical Genetics* 46(5-6): 329-346.

Tanabe, A.S. 2007. Kakusan: A computer program to automate the selection of a nucleotide substitution model and the configuration of a mixed model on multilocus data. *Molecular Ecology Notes* 7(6): 962-964.

Tucker, P.K., Sandstedt, S.A. & Lundrigan, B.L. 2005. Phylogenetic relationships in the subgenus *Mus* (genus *Mus*, family Muridae, subfamily Murinae): Examining gene trees and species trees. *Biological Journal of the Linnean Society* 84(3): 653-662.

Wang, S., Cong, H., Kong, L., Morokawa, M. & Li, Y. 2015. Complete mitochondrial genome of the greater bandicoot rat *Bandicota indica* (Rodentia: Muridae). *Mitochondrial DNA* 27(6): 4349-4350.

Yong, B., Wei, H., Jia, Q. & Chen, S. 2016. Sequencing and analysis of complete mitochondrial genome of *Niviventer fulvescens* (Muridae). *Mitochondrial DNA* 27(5): 3650-3651.

Yong, H.S. 1969. Karyotypes of Malayan rats (Rodentia, Muridae, genus *Rattus* Fischer). *Chromosoma* 27(3): 245-267.

Yong, H.S. 1968. *A Comparative Study of the Genetics and Systematics of the Malayan Species of* Rattus *Fischer*. PhD Thesis. Kuala Lumpur: University of Malaya (Unpublished).

Yong, H.S. & Eamsobhana, P. 2013. Definitive rodent hosts of the rat lungworm *Angiostrongylus cantonensis*. *Raffles Bulletin of Zoology* 29: 111-115.

Yong, H.S. & Dhaliwal, S.S. 1980. Immunoelectrophoretic serum-protein patterns of three commensal rats in Peninsular Malaysia. *Malaysian Journal of Science* 6: 5-7.

Yong, H.S., Dhaliwal, S.S. & Tee, K.L. 1972. Chromosome studies of three Malayan island rats. *Cytologia* 37(3): 507-511.

Zhao, F., Zhang, T., Su, J., Nevo, E. & Lin, G. 2014. Mitochondrial genome of bamboo rat *Rhizomys pruinosus*. *Mitochondrial DNA* 25(5): 381-382.

Zhu, D., Huang, J., Kang, C., Song, X., Yue, B. & Zhang, X. 2016. The complete mitochondrial genome of *Leopoldamys edwardsi* (Rodentia: Muridae). *Mitochondrial DNA* 27(3): 1882-1884.

Sze-Looi Song*
Institute for Advanced Studies
University of Malaya
50603 Kuala Lumpur, Federal Territory
Malaysia

Sze-Looi Song* & Phaik-Eem Lim
Institute of Ocean and Earth Sciences
University of Malaya
50603 Kuala Lumpur, Federal Territory
Malaysia

Hoi-Sen Yong
Institute of Biological Sciences
Faculty of Science
University of Malaya
50603 Kuala Lumpur, Federal Territory
Malaysia

Ji Tan
Department of Agricultural and Food Sciences
Universiti Tunku Abdul Rahman
31900 Kampar, Perak Darul Ridzuan
Malaysia

*Corresponding author; email: szelooi@um.edu.my

Received: 26 December 2019
Accepted: 22 September 2020