Tropical Biomedicine 28(2): 457–463 (2011)

Sequence analysis on the mitochondrial *COX*I gene of recent clinical isolates of *Plasmodium knowlesi* in Klang valley, peninsular Malaysia

Mun Yik Fong*, Yee-Ling Lau, Lit-Chien Chin and Abdulsalam M.Q. Al-Mekhlafi
Department of Parasitology, Faculty of Medicine, University of Malaya, 50603 Kuala Lumpur, Malaysia
* Corresponding author e-mail: fongmy@um.edu.my
Received 26 February 2011; received in revised form 18 April 2011; accepted 22 April 2011

Abstract. The cytochrome oxidase subunit I (*COX*I) gene sequences of three recent (2007-2008) clinical *Plasmodium knowlesi* isolates from Klang Valley, peninsular Malaysia, were determined and compared with those of older (1960's) peninsular Malaysia, recent isolates from Sarawak (on Borneo Island), and an isolate from Thailand. Multiple alignment of the sequences showed that the three clinical isolates were more similar to the older peninsular Malaysia isolates than to those from Sarawak and Thailand. Phylogenetic tree based on the *COX*I sequences revealed three distinct clusters of *P. knowlesi*. The first cluster consisted of isolates from peninsular Malaysia, the second consisted of Sarawak isolates and the third composed of the Thailand isolate. The findings of this study highlight the usefulness of mitochondrial *COX*I gene as a suitable marker for phylogeographic studies of *P. knowlesi*.

INTRODUCTION

Malaria is caused by protozoa of the genus *Plasmodium*. Four species are responsible for human malaria: *Plasmodium falciparum*, *Plasmodium vivax*, *Plasmodium malariae*, and *Plasmodium ovale*. Zoonotic infections by animal malaria parasites have initially been thought to be rare. However, in 2004, a large focus of human infection by *Plasmodium knowlesi*, a simian malaria parasite, was reported in Sarawak, on Borneo Island (Singh *et al.*, 2004). Recently, the Institute for Medical Research, Kuala Lumpur, reported 55.9% of human blood samples received from peninsular Malaysia between 2005 and 2008 were positive for *P. knowlesi* (Vythilingam *et al.*, 2008).

Despite the increasing importance of *P. knowlesi* as a zoonotic agent, relatively little is known regarding the population genetics of this parasite, especially of those in peninsular Malaysia. Vythilingam *et al.* (2008) studied clustering of *P. knowlesi* isolates from peninsular Malaysia based on the circumsporozoite (*csp*) gene sequence. However, *csp* codes for a sporozoite surface antigen which is subjected to strong selective immune pressure (De La Cruz *et al.*, 1989). Hence, variation in the *csp* sequence may not give a true reflection of genetic history of the parasite. The SSU rRNA (18S rRNA) is a genetic locus which is commonly used for genetic studies, but it is located in the nucleus and can be subjected to gene conversion and may not reflect the history of species and the times of divergence (Corredor & Enea, 1993).

Mitochondrial DNA (mtDNA) is suitable a genetic marker for phylogenetic studies because mitochondrial genes evolve more rapidly than nuclear genes. Furthermore, mtDNA does not undergo recombination, is maternally inherited, and has a simple sequence organization (Harrison, 1989). It has been shown in some studies that mtDNA haplotypes (unique mtDNA sequences) are

often geographically localized within the range of species (Avise *et al*., 1979; Phillips, 1994), thus making it possible to determine the geographic position in a phylogenetic tree (Morgan-Richards *et al*., 2001). Mitochondrial genes such as cytochrome oxidase subunit I (*COX*I), cytochrome oxidase subunit II (*CO*II) and cytochrome b (cytb) have been used for intraspecific phylogeographic studies of a variety of animal and insect species (Lunt *et al*., 1998; Caterino *et al*., 2000; Simmons & Weller, 2001; Zhou *et al*., 2011).

In the present study, we determined the *COX*I sequences of three *P. knowlesi* isolates from clinical cases which were admitted to the University of Malaya Medical Centre (UMMC) in July 2007 – June 2008 (Lee *et al*., 2010). The three patients had history of traveling into jungles or staying in areas with close proximity to jungles in the Klang Valley, which is located in the central part of Selangor, a state in west coast of peninsular Malaysia. The *COX*I sequences obtained were compared with those of older *P. knowlesi* isolates from mosquito (1960), monkey (1962) and human (1965); and more recent isolates from Sarawak (Lee *et al*., 2011) and Thailand (Jongwutiwes *et al*., 2004).

MATERIALS AND METHODS

Extraction of total DNA

Total DNA of the three *P. knowlesi* isolates (designated as PkUMMC1, PkUMMC5 and PkUMMC9) was extracted from the blood of the respective patients, using the QIAGEN Blood DNA Extraction kit (Germany). In each extraction, 100 µl of blood was used. The purified DNA was suspended to a final volume of 30 µl.

PCR, cloning and sequencing of the *P. knowlesi COX*I gene

Primers (forward: 5'-GACATTACCTGTA TTAAC-3'; reverse: 5'-GGTAAGAATG TTAAAACA-3') for the PCR of *P. knowlesi COX*I gene were designed based on the sequences of this gene in the GenBank nucleotide database. The PCR amplified a 692 bp region within the *COX*I gene. PCR was carried out in a 25 ml reaction mixture containing 10 mM Tris–HCl (pH 8.3), 2 mM MgCl2, 50 mM KCl, 0.01% gelatin, 200 mM of each deoxynucleoside triphosphate, 20 pmol of each primer, 1 U of *Taq* polymerase (Fermentas Life Sciences, Canada). The PCR mixture was pre-heated at 95°C for 10 min for initial denaturation before 30 cycles of amplification, which consisted of denaturation at 94°C for 1 min, annealing at 54°C for 1 min, and elongation at 72°C for 2 min. Final extension of the reaction was carried out at 72°C for 10 min. PCR fragment was cloned into plasmid pGEM-T, as according to the manufacturer (Promega Corporation, Madison, USA). Recombinant pGEM-T plasmids harbouring the cloned fragment were sent to a commercial laboratory for DNA sequencing.

Reference *Plasmodium* species *COX*I sequences

Reference *Plasmodium* species *COX*I sequences used for multiple sequence alignment and phylogenetic analysis were obtained from the GenBank database. The sequences included AY722797 (*P. knowlesi* mosquito isolate); AB444106 (*P. knowlesi* monkey isolate); AB444108 (*P. knowlesi* human isolate); EU880446 (*P. knowlesi* KH45); EU880454 (*P. knowlesi* KH176); EU880458 (*P. knowlesi* KH229); EU880465 (*P. knowlesi* KH381); EU880468 (*P. knowlesi* KH433); EU880474 (*P. knowlesi* LT4); EU880485 (*P. knowlesi* LT48); EU880476 (*P. knowlesi* LT15); EU880489 (*P. knowlesi* LT53); EU880498 (*P. knowlesi* LT57); AY598141 (*P. knowlesi* A1, Thailand); AY722799 (*P. fragile*); AY598035 (*P. vivax*); AB444131 (*P. cynomolgi*); AB444112 (*P. inui*); AB444133 (*P. fieldi*); AB354575 (*P. coatneyi*).

Sequence data analysis

Nucleotide sequences were aligned using the CLUSTAL-W programme which was available on-line (http://www.ebi.ac.uk/Tools/msa/clustalw2). Pairwise nucleotide differences and genetic distance (defined as p-distance) between the sequences were calculated using MEGA4 (Tamura *et al*., 2007). Phylogenetic tree was constructed

using the Maximum Parsimony (MP) method, also described in MEGA4. In the construction of the phylogenetic tree, bootstrap replicates of 1000 were used to test the robustness of the tree.

RESULTS

Alignment of the *COX*I sequences of PkUMMC1 (GenBank JF323868), PkUMMC5 (GenBank JF323869) and PkUMMC9 (GenBank JF323870) with those of other *P. knowlesi* isolates and *Plasmodium* species revealed that the gene was conserved in terms of size (full sequence alignment data not shown). All the *COX*I sequences were 692 bp in length, without any insertion/deletions (indels) or gaps. The summarized comparative alignment which shows the positions of nucleotide difference in the gene is presented in Figure 1. Using the *COX*I of PkUMMC1 as reference for comparison within *P. knowlesi*, five positions of nucleotide difference were noted: 76 (G↔A), 204 (C↔T), 295 (T↔C), 626 (G↔A) and 646 (C↔T). The alignment also revealed that isolates from peninsular Malaysia were more similar to each other than to those from Sarawak and Thailand. This is reflected in Table 1, which presents the number of pairwise nucleotide differences and genetic distance (p-distance) between the sequences. The number of nucleotide differences among the peninsular Malaysia isolates ranged from 0-2, which corresponded to genetic distance range of 0.000-0.003. When the peninsular Malaysia isolates were compared with isolates from Sarawak, number of nucleotide differences ranged from 1-4, corresponding to a relatively higher genetic distance range of 0.001-0.006. The Sarawak isolates were rather homogenous, with majority of them having identical gene sequences. The highest number of nucleotide differences was 1 (genetic distance = 0.001). The number of nucleotide differences between the peninsular Malaysia isolates and the Thailand isolate ranged from 1-3 (genetic distance range, 0.001-0.004).

The genetic relationships of the *COX*I sequences can be better inferred if presented in the form of a phylogenetic tree. In our study, the tree was constructed using the Maximun Parsimony approach, and *COX*I sequences of *Plasmodium fragile*, *P. vivax*,

```
                          0000000011111122222222333344444444444555555555555555666666
                          1477999912346601556679244801115677889001223334444677923447 8
                          2526036740815840232795725721466247062146281470369726369268 7
PkUMMC1 (P. M'sia)        ATTGTTTCCCATATCCATGATTTATCTTTGTTAAAAATTTTATTTTCTAAGGGGATCTC
PkUMMC5 (P. M'sia)        ...A.........T.............................................
PkUMMC9 (P. M'sia)        ...A.........T.............................................
Mosquito (P. M'sia, 1960) ...A.........T.............................................
Monkey (P. M'sia, 1962)   ...A.........T.............................................
Human (P. M'sia, 1965)    ...A.........T.............................................
Pk KH45 (Sarawak, Borneo) ...A.........T......C.....................................
Pk KH176 (Sarawak, Borneo)...A.........T......C.....................................
Pk KH229 (Sarawak, Borneo)...A.........T......C.....................................
Pk KH381 (Sarawak, Borneo)...A.........T......C.....................................
Pk KH433 (Sarawak, Borneo)...A.........T......C.....................................A.....
Pk LT4 (Sarawak, Borneo)  ...A.........T......C.....................................
Pk LT48 (Sarawak, Borneo) ...A.........T......C.....................................
Pk LT15 (Sarawak, Borneo) ...A.........T......C.....................................
Pk LT53 (Sarawak, Borneo) ...A.........T......C.....................................
Pk LT57 (Sarawak, Borneo) ...A.........T......C.....................................
Pk A1 (Thailand)          ...A.........T.....................................T..
Plasmodium fragile        ...A...TTT.CT.TT...GC....T.........T...C.C..CCT..TC.....T.T
Plasmodium vivax          T..A...TTTG.T.TT.........T.........T....C.C..CT…….....ATA.
Plasmodium cynomolgi      C..A.A.TTT..T.TT.........T.........T....C....CT…..C....AT..
Plasmodium inui           TC.A...TTT..TGTTT.A....CCT...A.C.TTT..A.C.AC.CTC.T..A..AT..
Plasmodium fieldi         C.CA..CTTT..T.TT.C....C..T..C..A...T.........CT..T.A...AT..
Plasmodium coatneyi       ...AC..TTT...CTT.........TCC..C.GTCTTC....C..C..TTCA..T.T..
```

Figure 1. Comparative alignment showing the positions (denoted by the vertical numbers) of nucleotide changes in the *COX*I sequences of *Plasmodium knowlesi* isolates and other *Plasmodium* species. Dots indicate nucleotide identical to those of PkUMMC1

Table 1. Pairwise comparisons of nucleotide difference and genetic difference (p-distance) among *Plasmodium knowlesi* isolates from peninsular Malaysia, Sarawak and Thailand. The figures above the diagonal black boxes are the genetic differences

	PkUMMC1 (P. M'sia,)	PkUMMC5 (P. M'sia,)	PkUMMC9 (P. M'sia,)	Mosquito isolate (P. M'sia, 1960)	Monkey isolate (P. M'sia, 1962)	Human isolate (P. M'sia, 1965)	Pk KH45 (Sarawak, Borneo)	Pk KH176 (Sarawak, Borneo)	Pk KH229 (Sarawak, Borneo)	Pk KH381 (Sarawak, Borneo)	Pk KH433 (Sarawak, Borneo)	Pk LT4 (Sarawak, Borneo)	Pk LT48 (Sarawak, Borneo)	Pk LT15 (Sarawak, Borneo)	Pk LT53 (Sarawak, Borneo)	Pk LT57 (Sarawak, Borneo)	Pk A1 (Thailand)
PkUMMC1 (P. M'sia)		0.003	0.001	0.001	0.001	0.003	0.004	0.004	0.004	0.004	0.006	0.004	0.004	0.004	0.004	0.004	0.004
PkUMMC5 (P. M'sia)	2		0.001	0.001	0.001	0.000	0.001	0.001	0.001	0.001	0.003	0.001	0.001	0.001	0.001	0.001	0.001
PkUMMC9 (P. M'sia)	1	1		0.000	0.000	0.001	0.003	0.003	0.003	0.003	0.004	0.003	0.003	0.003	0.003	0.003	0.003
Mosquito isolate (P. M'sia, 1960)	1	1	0		0.000	0.001	0.003	0.003	0.003	0.003	0.004	0.003	0.003	0.003	0.003	0.003	0.003
Monkey isolate (P. M'sia, 1962)	1	1	0	0		0.001	0.003	0.003	0.003	0.003	0.004	0.003	0.003	0.003	0.003	0.003	0.003
Human isolate (P. M'sia, 1965)	2	0	1	1	1		0.001	0.001	0.001	0.001	0.003	0.001	0.001	0.001	0.001	0.001	0.001
Pk KH45 (Sarawak, Borneo)	3	1	2	2	2	1		0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.003
Pk KH176 (Sarawak, Borneo)	3	1	2	2	2	1	0		0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.003
Pk KH229 (Sarawak, Borneo)	3	1	2	2	2	1	0	0		0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.003
Pk KH381 (Sarawak, Borneo)	3	1	2	2	2	1	0	0	0		0.001	0.000	0.000	0.000	0.000	0.000	0.003
Pk KH433 (Sarawak, Borneo)	4	2	3	3	3	2	1	1	1	1		0.001	0.001	0.001	0.001	0.001	0.004
Pk LT4 (Sarawak, Borneo)	3	1	2	2	2	1	0	0	0	0	1		0.000	0.000	0.000	0.000	0.003
Pk LT48 (Sarawak, Borneo)	3	1	2	2	2	1	0	0	0	0	1	0		0.000	0.000	0.000	0.003
Pk LT15 (Sarawak, Borneo)	3	1	2	2	2	1	0	0	0	0	1	0	0		0.000	0.000	0.003
Pk LT53 (Sarawak, Borneo)	3	1	2	2	2	1	0	0	0	0	1	0	0	0		0.000	0.003
Pk LT57 (Sarawak, Borneo)	3	1	2	2	2	1	0	0	0	0	1	0	0	0	0		0.003
Pk A1 (Thailand)	3	1	2	2	2	1	2	2	2	2	3	2	2	2	2	2	

Plasmodium cynomolgi, *Plasmodium inui*, *Plasmodium fieldi*, and *Plasmodium coatneyi* were included as outgroups. In the phylogenetic tree (Figure 2), three distinct clusters of *P. knowlesi* were observed. The first consisted of isolates from peninsular Malaysia, the second consisted of Sarawak isolates, and the third composed of the Thailand isolate. Interestingly, the peninsular Malaysia cluster was divided into three subgroups, which basically followed the nucleotide differences (or genetic distance) of the isolates. For example, PkUMM5 and human isolate (1965) had identical nucleotide sequences, thus they were grouped together. Similarly, the mosquito isolate (1960), monkey isolate (1962) and PkUMMC9 formed another subgroup since their sequences were identical. These two subgroups differed by only one nucleotide, at position 204 (C↔T). The third subgroup consisted of isolate PkUMMC1, and its nucleotide sequence differed from the first and second subgroups at positions 76 (G↔A) and 204 (CT), respectively.

DISCUSSION

This study was conducted to compare the sequence variation of the *COX*I gene between three recent (2007-2008) clinical isolates from Klang Valley, and also to compare their sequences with those of some other *P. knowlesi* isolates which were available in GenBank. In these comparisons, isolates which were separated temporally and geographically from these three *P. knowlesi* isolates were selected. For the purpose of temporal comparison, isolates of the 1960's from different origins were chosen, namely mosquito, monkey and human. For geographical comparison, recent isolates from Sarawak, which is separated from Peninsular Malaysia by the South China Sea were selected. A clinical isolate from Thailand was also included in the analysis. The genetic relationship of all the isolates was determined in a phylogenetic tree.

The advantage of using *COX*I in the study lies in the consistency of length of the gene. No gaps or insertions were found within this

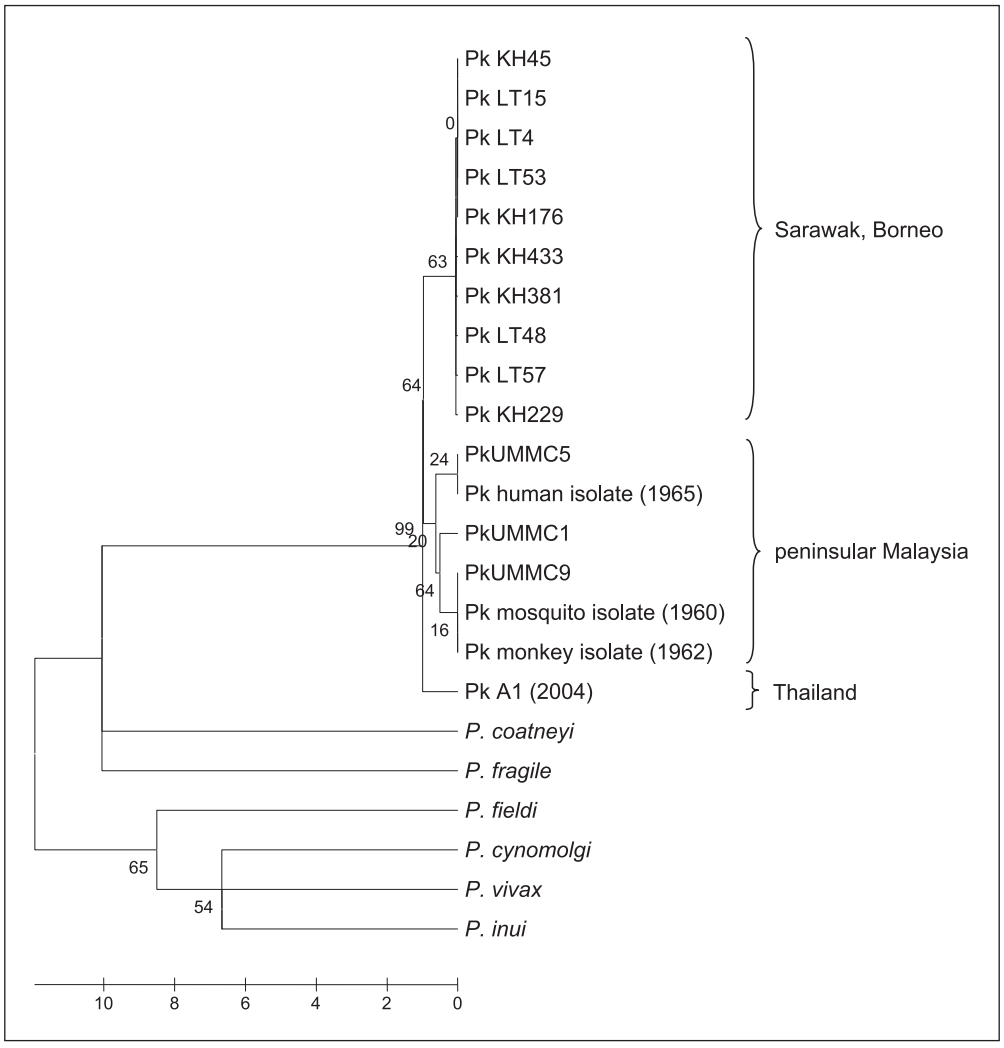

Figure 2. Phylogenetic tree based on *COX*I gene sequences of *Plasmodium knowlesi* (Pk) isolates and other *Plasmodium* species. The tree is constructed using Maximum Parsimony method. The percentage of replicate trees in which the associated isolates cluster together in the bootstrap test (1000 replicates) are shown next to the branches. The phylogenetic analysis was conducted using MEGA4 (Tamura *et al.*, 2007)

gene of *P. knowlesi* isolates or strains. No sequence editing or gap/insertion weightage had to be taken into consideration in the multiple sequence alignment and subsequent phylogenetic analysis. This is, however, not possible if the *csp* gene or SSU rRNA is used. The *csp* gene is very variable in length. Tandem repeats of different sizes and nucleotide sequences are located in the centre region, whereas the 5' and 3' coding and noncoding sequences flanking the repeats are conserved (Sharma *et al.*, 1985). For instance, Vythilingam *et al.* (2008) reported the *csp* genes of peninsular Malaysia isolates ranged from 1050 to1128 bp. The SSU rRNA sequence length variation in *P. knowlesi* has also been observed (Singh *et al.*, 2004).

This study also highlights the high resolution power of using *COX*I sequence and Maximum Parsimony approach. Although the number of nucleotide differences between the peninsular Malaysia, Sarawak and Thailand isolates was small (1-4 of 692 nucleotides), the approach was able to distinctly separate these isolates on the basis of geographical origin. The separation

between the peninsular Malaysia and Sarawak isolates is attributed to the single nucleotide difference at position 295 (fixed mutation) within the *COX*I gene. However, in the phylogenetic trees of Vythilingam *et al.* (2008) based on *csp* sequences, some peninsular Malaysia isolates were grouped together with isolates from Sarawak. Similarly, the phylogenetic trees based on SSU rRNA and *csp* sequences obtained by Singh *et al.* (2004) showed cluster containing both peninsular Malaysia (Nuri strain) and Sarawak isolates.

Temporal separation among the peninsular Malaysia isolates was not distinct. No major sequence differences were seen despite some of them being separated by time for nearly 50 years (1960-2008). In fact, some of the old isolates had identical sequence with the recent clinical isolates [human isolate (1965) with PKUMMC5; mosquito isolate (1960) and monkey isolate (1962) with PkUMMC9].

In conclusion, the findings of our simple study demonstrate the usefulness of mitochondrial *COX*I gene as a suitable marker for phylogeographic studies of *P. knowlesi*, which in turn can help in tracing movement of strains across geographical regions. For future work, isolates from other countries in Southeast Asian will be included in order to obtain a more comprehensive picture of the phylogeographic distribution of *P. knowlesi* strains in the region.

Acknowledgement. This study was supported by the University of Malaya Research Grant RG021/09HTM.

REFERENCES

Avise, J.C., Giblin-Davidson, C., Laerm, J., Patton, J.C. & Lansman, R.A. (1979). Mitochondrial DNA clones and matriarchal phylogeny within and among geographic populations of the pocket gopher, *Geomys pinetis*. *Proceedings of the National Academy of Sciences USA* **76**: 6694-6698.

Caterino, M.S., Cho, S. & Sperling, F.A. (2000). The current state of insect molecular systematics: a thriving Tower of Babel. *Annual Review of Entomology* **45**: 1-54.

Corredor, V. & Enea, V. (1993). Plasmodial ribosomal RNA as phylogenetic probe: a cautionary note. *Molecular Biology and Evolution* **10**: 924-926.

De La Cruz, V.F., Maloy, W.L., Miller, L.H., Good, M.F. & McCutchan, T.F. (1989). The immunologic significance of variation within malaria circumsporozoite protein sequences. *Journal of Immunology* **142**: 3568-3575.

Harrison, R.G. (1989). Animal mitochondrial DNA as a genetic marker in population and evolutionary biology. *Trends in Ecology and Evolutions* **4**: 6-11.

Jongwutiwes, S., Putaporntip, C., Iwasaki, T., Sata, T. & Kanbara, H. (2004). Naturally acquired *Plasmodium knowlesi* malaria in human, Thailand. *Emerging Infectious Diseases* **10**: 2211-2213.

Lee, C.E., Adeeba, K. & Freigang, G. (2010). Human *Plasmodium knowlesi* infections in Klang Valley, Peninsula Malaysia: a case series. *Medical Journal of Malaysia* **65**: 63-65.

Lee, K.S., Divis, P.C.S., Zakaria, S.K., Matusop, A., Julin, R.A., Conway, D.J., Cox-Singh, J. & Singh, B. (2011). *Plasmodium knowlesi*: Reservoir Hosts and Tracking the Emergence in Humans and Macaques. *PLoS Pathogen* **7**(4): e1002015. doi:10.1371/journal.ppat. 1002015

Lunt, D.H., Ibrahim, K.M. & Hewitt, G.M. (1998). mtDNA phylogeography and postglacial patterns of subdivision in the meadow grasshopper *Chorthippus parallelus*. *Heredity* **80**: 633-641.

Morgan-Richards, M., Trewick, S.A. & Wallis, G.P. (2001). Chromosome races with Pliocene origins: evidence from mtDNA. *Heredity* **86**: 303-312.

Phillips, C.A. (1994). Geographic distribution of mitochondrial DNA variants and the historical biogeography of the spotted salamander, *Ambystoma maculatum*. *Evolution* **48**: 597-607.

Sharma, S., Svec, P., Mitchell, G.H. & Godson, G.N. (1985). Diversity of circumsporozoite antigen genes from two strains of the malarial parasite *Plasmodium knowlesi*. *Science* **229**: 779-782.

Simmons, R.B. & Weller, S.J. (2001). Utility and evolution of cytochrome b in insects. *Molecular Phylogenetics and Evolution* **20**: 196-210.

Singh, B., Lee, K.S., Matusop, A., Radhakrishnan, A., Shamsul, S.S., Cox-Singh, J., Thomas, A. & Conway, D.J. (2004). A large focus of naturally acquired *Plasmodium knowlesi* infections in human beings. *Lancet* **363**: 1017-1024.

Tamura, K., Dudley, J., Nei, M. & Kumar, S. (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular Biology and Evolution* **24**: 1596-1599.

Vythilingam, I., Noorazian, Y.M., Huat, T.C., Jiram, A.I., Yusri, Y.M., Azahari, A.H., Norparina, I., Noorrain, A., Lokmanhakim, S. (2008). *Plasmodium knowlesi* in humans, macaques and mosquitoes in peninsular Malaysia. *Parasites and Vectors* **1**: 26.

Zhou, C., Li, M., Yuan, K., Hu, N. & Peng, W. (2011). Phylogeography of *Ascaris lumbricoides* and *A. suum* from China. *Parasitology Research* DOI 10.1007/s00436-011-2260-4 [Epub ahead of print].