Sains Malaysiana 51(8)(2022): 2435-2447
http://doi.org/10.17576/jsm-2022-5108-08

Characterisation of *Tetrastigma rafflesiae* Mitochondrial Genes and Assessment of their Potential as Sequence Markers

(Pencirian Gen Mitokondria *Tetrastigma rafflesiae* dan Penilaian Potensi Gen Tersebut sebagai Penanda Jujukan)

MOHAMAD HAFIZZUDIN-FEDELI[1], MOHD-FAIZAL ABU-BAKAR[2], MOHD-NOOR MAT-ISA[1,2], A. LATIFF[1], MOHD FIRDAUS-RAIH[3,4] & KIEW-LIAN WAN[1,*]

[1]Department of Biological Sciences and Biotechnology, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor Darul Ehsan, Malaysia
[2]Malaysia Genome and Vaccine Institute, National Institutes of Biotechnology Malaysia, 43000 Kajang, Selangor Darul Ehsan, Malaysia
[3]Department of Applied Physics, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor Darul Ehsan, Malaysia
[4]Institute of Systems Biology, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor Darul Ehsan, Malaysia

Received: 14 June 2021/Accepted: 10 February 2022

ABSTRACT

Tetrastigma rafflesiae (Miq.) Planch. is a climbing plant species that is known for its unique relationship with holoparasitic plants of Rafflesiaceae. Knowledge on the mitochondrial genes of this species may contribute towards the development of molecular approaches for species identification. This study aimes to identify and characterise genes from the *T. rafflesiae* mitochondrial genome (mitogenome) and assess their potential to discriminate different *Tetrastigma* species. Mitochondrial-specific sequences were first selected by mapping *T. rafflesiae* whole-genome sequences to mitogenomes from several reference plant species. *De novo* assembly of these selected sequences produced a *T. rafflesiae* mitogenome with a size of 336 kb. Gene annotation revealed that the *T. rafflesiae* mitogenome contains at least 40 protein coding genes, 20 tRNAs and two rRNAs. Phylogenetic analysis using several mitochondrial genes, namely *ccmB, cob, matR, nad6* and *rps3* was able to differentiate *T. rafflesiae* from three other *Tetrastigma* species, indicating the potential of these genes as species-specific sequence markers. These findings supplement additional genetic information on *T. rafflesiae* and may aid in the effort of species classification and conservation.

Keywords: Molecular markers; plant mitogenome; phylogeny; species identification

ABSTRAK

Tetrastigma rafflesiae (Miq.) Planch. adalah spesies tumbuhan memanjat yang terkenal kerana hubungannya yang unik dengan tumbuhan holoparasit daripada Rafflesiaceae. Pengetahuan mengenai gen mitokondria spesies ini dapat menyumbang ke arah pembangunan pendekatan molekul untuk pengecaman spesies. Kajian ini bertujuan untuk mengenal pasti dan mencirikan gen daripada genom mitokondria (mitogenom) *T. rafflesiae* dan menilai potensi gen tersebut dalam membezakan spesies *Tetrastigma* yang berbeza. Jujukan khusus mitokondria pada mulanya dipilih melalui pemetaan jujukan keseluruhan genom *T. rafflesiae* kepada mitogenom daripada beberapa spesies tumbuhan rujukan. Penghimpunan *de novo* jujukan terpilih ini menghasilkan mitogenom *T. rafflesiae* yang bersaiz 336 kb. Anotasi gen menunjukkan bahawa mitogenom *T. rafflesiae* mengandungi sekurang-kurangnya 40 gen pengekodan protein, 20 tRNA dan dua rRNA. Analisis filogenetik menggunakan beberapa gen mitokondria iaitu *ccmB*, *cob*, *matR*, *nad6* dan *rps3* berupaya membezakan *T. rafflesiae* daripada tiga spesies *Tetrastigma* yang lain lalu menunjukkan potensi kesemua gen ini sebagai penanda jujukan yang khusus bagi spesies. Penemuan ini menyumbang maklumat genetik tambahan mengenai *T. rafflesiae* dan boleh membantu usaha pengkelasan dan pemuliharaan spesies.

Kata kunci: Filogeni; mitogenom tumbuhan; penanda molekul; pengecaman spesies

INTRODUCTION

Tetrastigma (Miq.) Planch. is a genus of climbing plant belonging to Vitaceae that consists of about 95 species distributed throughout the Asian-Oceania tropics and subtropics from Sino-Himalaya to Taiwan and Malesia, extending to Oceania regions (Chen et al. 2011). Several *Tetrastigma* species have been known to exhibit

medicinal properties. For instance, *T. hemsleyanum* has been used traditionally by the locals in China to remedy inflammation and fever, as well as an analgesic and to improve blood circulation (Fu et al. 2011; Ji et al. 2021; Liu et al. 2002). *Tetrastigma* species also attract a major interest due to their exclusive role as host plant to the parasitic plants in Rafflesiaceae that includes all three genera, *Rafflesia* R. Br., *Rhinzanthes* Dumort and *Sapria* Griffith (Chen et al. 2011; Molina et al. 2014; Nikolov et al. 2014). These genera, which can produce a single flower for reproduction, are leafless, stemless and rootless thus making them totally reliant on their host for nutrients. The genus *Rafflesia* is the most well-known due to the gigantic reddish flower with a diameter of up to 100 cm that it can produce (Nais 2001). This genus is endemic to the tropical rainforests of Malaysia, Brunei, Kalimantan, Java, Sumatera, the Philippines and Southern Thailand.

One of the species of *Tetrastigma*, *T. rafflesiae*, has been known to host many *Rafflesia* species found in Malaysia and also in its distribution range. The plant can be found thriving in lowland dipterocarp forests, secondary forests, rocky ridges, limestones and hill dipterocarp forests in Peninsular Malaysia (Latiff 1983), while in Borneo Island, it can be found mostly in mixed dipterocarp forests and montane forests (Wan Zakaria et al. 2016). Interestingly, *T. rafflesiae* has also been reported to host most of the *Rafflesia* species that can be found in Indonesia, Brunei, the Philippines and Southern Thailand. Previous studies have recorded more than ten species of *Rafflesia* parasitising *T. rafflesiae*, including those in Peninsular Malaysia (Adam et al. 2016; Nais 2001).

Molecular phylogenetics has been used to study the evolutionary history of living organisms, population genetics, phylogeography and the evolution of genes and genomes (Paterson et al. 2010; Yang & Rannala 2012). Several genomic regions could be utilised for phylogenetic reconstruction and DNA barcoding-based identification (Christin et al. 2012; Wendel & Doyle 1998). In plants, nuclear ribosomal internal transcribed spacers (ITS) which are the internal transcribed spacers of the large subunit of ribosomal DNA, have been utilised as the sequence of choice for phylogenetic markers (Álvarez & Wendel 2003; Feliner & Rosselló 2007). More recently, plastid genes, including those from chloroplast genomes (plastomes), are being extensively used in the reconstruction of phylogenies to identify plant species (Hollingsworth et al. 2009; Li et al. 2015; Sun et al. 2016). However, unlike nuclear and chloroplast genes, mitochondrial sequences are rarely used to reconstruct phylogenies or for DNA barcoding in higher plants (Donnelly et al. 2017; Govindarajulu et al. 2015). This is due to several limitations of plant mitochondrial genomes (mitogenomes), including extremely slow nucleotide substitution rate (Aguileta et al. 2014; Nabholz et al. 2009), being maternally inherited (Corriveau & Coleman 1988; Reboud & Zeyl 1994) and difficulties in assembling complete plant mitogenomes due to their complex and variable structure (Fauron et al. 2004; Smith & Keeling 2015). However, despite these challenges, comparative mitogenome studies are still relevant to show the potential evolutionary history and structural dynamics among certain plant species (Duminil 2014). This was demonstrated by previous studies that employed several mitochondrial sequences to resolve phylogeny among parasitic angiosperms (Barkman et al. 2007) and the *Pinus* genus (Donnelly et al. 2017).

In this study, we have identified mitochondrial specific reads of *T. rafflesiae* by mapping whole genome sequencing reads to reference plant mitogenomes and assembled the reads *de novo*. This was followed by the characterisation of the mitogenome gene content through annotation and the comparison with other reference plant mitogenomes. Several protein coding genes were identified and assessed for their potential as sequence marker to differentiate *T. rafflesiae* from other *Tetrastigma* species. This understanding could pave the way to develop a strategy of using mitochondrial based markers for *Tetrastigma* identification and classification.

Methods

de novo Assembly of *T. rafflesiae* Mitogenome

To assemble the *T. rafflesiae* mitogenome, we first identified mitochondrial specific reads from the Illumina whole genome sequence dataset generated from tissue samples of *T. rafflesiae* grown in the forest of the Malaysia Genome Institute (MGI). This was carried out by Bowtie2 (v2.4.1) (Langmead & Salzberg 2012) that mapped the pre-processed whole genome reads to mitogenome sequences of reference plant species. Twenty-two flowering plant species mitogenomes with size ranging from 119 kb in *Amborella trichopoda* to 1.3 Mb in *Hevea brasiliensis* were used as reference mitogenomes (Table 1). All the mitogenome sequences were retrieved and downloaded from the GenBank database in FASTA format. Sequences mapped by Bowtie2 were then processed by Velvet (v1.2.10) (Zerbino & Birney 2008) that assembled the short reads. The assembled mitogenome of *T. rafflesiae* was then

analysed using SSPACE (v2.1.1) (Boetzer et al. 2010) and MEDUSA (v1.6) (Bosi et al. 2015) to increase the scaffold length through scaffolding and contig extension processes and therefore, reduce the overall scaffold number. Closure of the gaps produced during the scaffolding process was performed using GapCloser (v1.0.1) to improve the overall contiguity and completeness of the assembled genome.

TABLE 1. Reference plant species used in mapping of *T. rafflesiae* mitogenome sequence reads

Species	Mitogenome size (bp)	GenBank Accession No.
Amborella trichopoda	118,716	KF754799.1
Arabidopsis thaliana	336,924	NC_001284.2
Beta vulgaris	368,801	NC_002511.2
Brassica napus	223,853	NC_008285.1
Carica papaya	476,890	NC_012116.1
Cucurbita pepo	982,833	NC_014050.1
Daucus carota	281,132	NC_017855.1
Gossypium hirsutum	621,884	NC_027406.1
Helianthus annuus	300,945	NC_023337.1
Hevea brasiliensis	1,325,823	AP014526.1
Liriodendron tulipifera	553,721	NC_021152.1
Malus domestica	396,947	NC_018554.1
Marchantia paleacea	186,609	NC_001660.1
Nicotiana tabacum	430,597	NC_006581.1
Oryza sativa	490,520	NC_011033.1
Phoenix dactylifera	715,001	NC_016740.1
Raphanus sativus	258,426	NC_018551.1
Ricinus communis	502,773	HQ874649.1
Triticum aestivum	452,528	AP008982.1
Vigna radiata	401,262	NC_015121.1
Vitis vinifera	773,279	NC_012119.1
Zea mays	596,630	NC_007982.1

ANNOTATION OF *T. Rafflesiae* MITOCHONDRIAL GENES

Gene annotation of the *T. rafflesiae* mitogenome was carried out using Mitofy (v1.3.1) (Alverson et al. 2010) that was specifically designed for the annotation of seed plant mitochondrial genome. Within Mitofy, BLASTX was used to search for the assembled *T. rafflesiae* mitogenome against the 41 different protein coding genes known in

seed plant mitogenome. BLASTN was used to search for matching query sequences against the 27 tRNA and three rRNA genes. BLASTX and BLASTN parameters used included maximum BLAST expect value for both protein genes and RNA genes set to 1e-3, and minimum percent identity set to 70. tRNAscan-SE (v2.11) was used to search for organellar tRNAs *de novo*. The annotated mitochondrial gene sequences in Mitofy were then verified through functional analyses in Blast2GO using a Gene Ontology (GO) annotation workflow that involved the alignment of each annotated gene sequence using BLASTX against the GenBank protein database and InterProScan search against the protein domain database that identified biological function of the gene based on protein domains and families. Then, the programme matched the ID of the annotated genes against the GO annotation database through a mapping process before assigning the most appropriate GO terms for each gene sequence (Götz et al. 2008). The output from Blast2GO analysis was used in WEGO (v2.0) (Ye et al. 2018, 2006) to generate the GO graph. Subsequently, the annotated protein coding gene sequences of *T. rafflesiae* were compared with those of *Arabidopsis thaliana*, a model plant that has been extensively studied, and *Vitis vinifera*, a species within the Vitaceae family.

COMPARATIVE ANALYSIS OF *T. rafflesiae* MITOCHONDRIAL GENES

The selection of gene candidates for comparative analysis were based on their full length, as determined from the mitogenome analysis, and the availability of their homologous sequences in other *Tetrastigma* species in the public database. In this study, five *T. rafflesiae* mitochondrial genes that met the criteria namely *ccmB, cob, matR*, *nad6,* and *rps3* were selected and analysed together with their homologues from three other *Tetrastigma* species which were *T. rumicispermum*, *T. voinierianum,* and *T. obovatum* (Chen et al. 2011), with *V. vinifera* (Goremykin et al. 2009) as an outgroup. All the reference gene sequences were retrieved and downloaded from GenBank. The multiple sequence alignment (MSA) analysis for each gene was performed using Clustal Omega (v1.2.4) (Sievers & Higgins 2014) and the results were used in subsequent phylogenetic tree reconstruction using MEGA X (v10.2) (Kumar et al. 2018). Maximum Likelihood (ML) statistical model and the General Time Reversible (GTR) nucleotide substitution were selected as the evolutionary model to construct the phylogenetic tree of the combined mitochondrial gene sequences from each species with 1000 bootstrap replication (Felsenstein 1981).

RESULTS

MAPPING AND THE ASSEMBLY OF *T. rafflesiae* MITOGENOME

Pre-processing of *T. rafflesiae* whole genome raw reads produced around 432 million high quality reads. Out of these, 275,984 reads mapped to the reference plant mitogenome sequences. The mapped reads were then assembled *de novo* using Velvet, and subsequently scaffolded using SSPACE, MEDUSA and GapCloser, producing a *T. rafflesiae* mitogenome assembly of 336 kb in length. There are 33 scaffold sequences within the assembly with an average length of 10 kb and the longest scaffold being 69 kb in length (Table 2).

TABLE 2. Assembly and scaffolding statistics of *T. rafflesiae* mitogenome

	Assembly	Scaffolding		
	Velvet	SSPACE	MEDUSA	GapCloser
Total number of contigs/scaffolds	314	248	33	33
Total contig/scaffold length (bp)	337,026	357,340	367,040	336,355
Contig/Scaffold N50 (bp)	2,215	3,142	19,389	18,700
Longest contig/scaffold (bp)	10,004	29,679	74,673	69,131
Average contig/scaffold length (bp)	1,073	1,441	11,122	10,193
Contig/Scaffold GC (%)	36.91	34.81	33.89	40.09

CHARACTERISATION OF GENES IN *T. rafflesiae* MITOGENOME

Based on the Mitofy results, the *T. rafflesiae* mitogenome was shown to consist of 40 protein coding mitochondrial, 20 tRNA and two rRNA genes (Table 3). The nucleotide sequences of protein coding genes in *T. rafflesiae* mitogenome were compared with the mitogenome of two other plant species namely *A. thaliana* and *V. vinifera* (Table 3). The results showed that *T. rafflesiae* and *V. vinifera* mitogenomes have the same number of protein coding genes (40 genes) while *A. thaliana* mitogenome has only 34 genes, despite *V. vinifera* being the largest (773 kb), compared with the mitogenomes of *T. rafflesiae* (336 kb) and *A. thaliana* (367 kb). This suggested that mitogenome size does not correlate with the number of genes in the plant mitochondria as the larger mitogenome size may be contributed by non-coding sequences including intron, intergenic region and foreign sequences (Alverson et al. 2010; Gualberto et al. 2014). The results also showed that the lengths of the protein coding sequences were comparable between the three species, and this information is essential in the determination of gene sequence completeness in the annotated mitogenome.

TABLE 3. Analysis of genes in *T. rafflesiae* mitogenome

Gene	Description	Length of protein coding genes (bp)		
		T. rafflesiae	*A. thaliana*	*V. vinifera*
Energy production				
Complex I				
nad1	NADH dehydrogenase subunit 1	996	504	714
nad2	NADH dehydrogenase subunit 2	1494	1815	1773
nad3	NADH dehydrogenase subunit 3	474	477	477
nad4	NADH dehydrogenase subunit 4	1500	579	1254
nad4L	NADH dehydrogenase subunit 4L	303	420	420
nad5	NADH dehydrogenase subunit 5	1983	1680	1974
nad6	NADH dehydrogenase subunit 6	618	735	810
nad7	NADH dehydrogenase subunit 7	1227	1611	1611
nad9	NADH dehydrogenase subunit 9	615	690	690
Complex II				
sdh3	succinate dehydrogenase subunit 3	282	n/a	435
sdh4	succinate dehydrogenase subunit 4	426	504	513
Complex III				
cob	apocytochrome b	1239	1299	1293
Complex IV				
cox1	cytochrome c oxidase subunit 1	1638	1698	1698
cox2	cytochrome c oxidase subunit 2	438	1020	555
cox3	cytochrome c oxidase subunit 3	855	915	915
Complex V				
atp1	ATP synthase F1 subunit 1	1563	1518	1569
atp4	ATP synthase F1 subunit 4	487	678	714
atp6	ATP synthase F1 subunit 6	744	1275	924
atp8	ATP synthase F1 subunit 8	483	591	597
atp9	ATP synthase F1 subunit 9	285	375	357
Cytochrome *c* biogenesis				
ccmB	cytochrome c biogenesis protein subunit B	586	738	666
ccmC	cytochrome c biogenesis protein subunit C	754	888	870
ccmFc	cytochrome c biogenesis protein subunit Fc	840	717	891
ccmFn	cytochrome c biogenesis protein subunit Fn	1782	1194	1905
Transcription				
matR	maturase	1965	1998	2019
Translation				

rps1	ribosomal protein S1	475	n/a	639
rps2	ribosomal protein S2	141	n/a	240
rps3	ribosomal protein S3	1620	2193	1893
rps4	ribosomal protein S4	1089	1206	1206
rps7	ribosomal protein S7	492	564	564
rps10	ribosomal protein S10	273	n/a	573
rps11	ribosomal protein S11	n/a	n/a	n/a
rps12	ribosomal protein S12	166	495	495
rps13	ribosomal protein S13	351	n/a	441
rps14	ribosomal protein S14	303	318	420
rps19	ribosomal protein S19	126	n/a	402
rpl2	ribosomal protein L2	888	1309	1041
rpl5	ribosomal protein L5	591	675	672
rpl10	ribosomal protein L10	60	n/a	594
rpl16	ribosomal protein L16	546	339	657
Transporter protein				
mttB	transporter protein	840	978	894
Transfer RNA				
Arg	Arginine	*	-	*
Asn	Asparagine	*	*	*
Asn-cp	Asparagine-cp	*	*	*
Asp	Aspartic acid	-	*	*
Cys-bacteria	Cysteine-bacteria	-	-	-
Cys-cp	Cysteine-cp	-	-	-
Cys-mt	Cysteine-mt	*	*	*
Gln	Glutamine	*	*	*
Glu	Glutamic acid	-	*	*
Gly	Glycine	*	*	*
His-cp	Histidine-cp	*	*	*
Ile	Isoleucine	*	*	*
Ile-cp	Isoleucine-cp	*	*	*
Leu	Leucine	-	-	-
Leu-cp	Leucine-cp	*	-	-
Lys	Lysine	*	*	*
Met-cp	Methionine-cp	*	*	*
Met-f	Methionine-f	-	*	*
Phe	Phenylalanine	*	*	*
Phe-cp	Phenylalanine-cp	-	-	-
Pro	Proline	*	*	*
Pro-cp	Proline-cp	*	-	*
Ser	Serine	*	*	*
Ser-cp	Serine-cp	*	*	-
Trp-cp	Tryptophan-cp	*	*	*
Tyr	Tyrosine	*	*	*
Val-cp	Valine-cp	*	-	-
Ribosomal RNA				
rrn5	5S ribosomal RNA	-	*	*
rrnL	16S ribosomal RNA	*	*	*
rrnS	18S ribosomal RNA	*	*	*

n/a: not available, (*) denotes present, (-) denotes absent

The annotated mitochondrial genes of *T. rafflesiae* were further analysed using Blast2GO to confirm the sequence identity through functional analysis. Based on the analysis results, 39 out of 40 annotated mitochondrial genes have matches with entries in Blast2GO. The GO analysis results of the annotated gene were plotted and categorised into cellular component, molecular function and biological process (Figure 1). In general, the GO annotation of the mitochondrial genes of *T. rafflesiae* showed that they are involved in biological and metabolic processes that are similar to the mitochondrial genes of other plant mitogenomes.

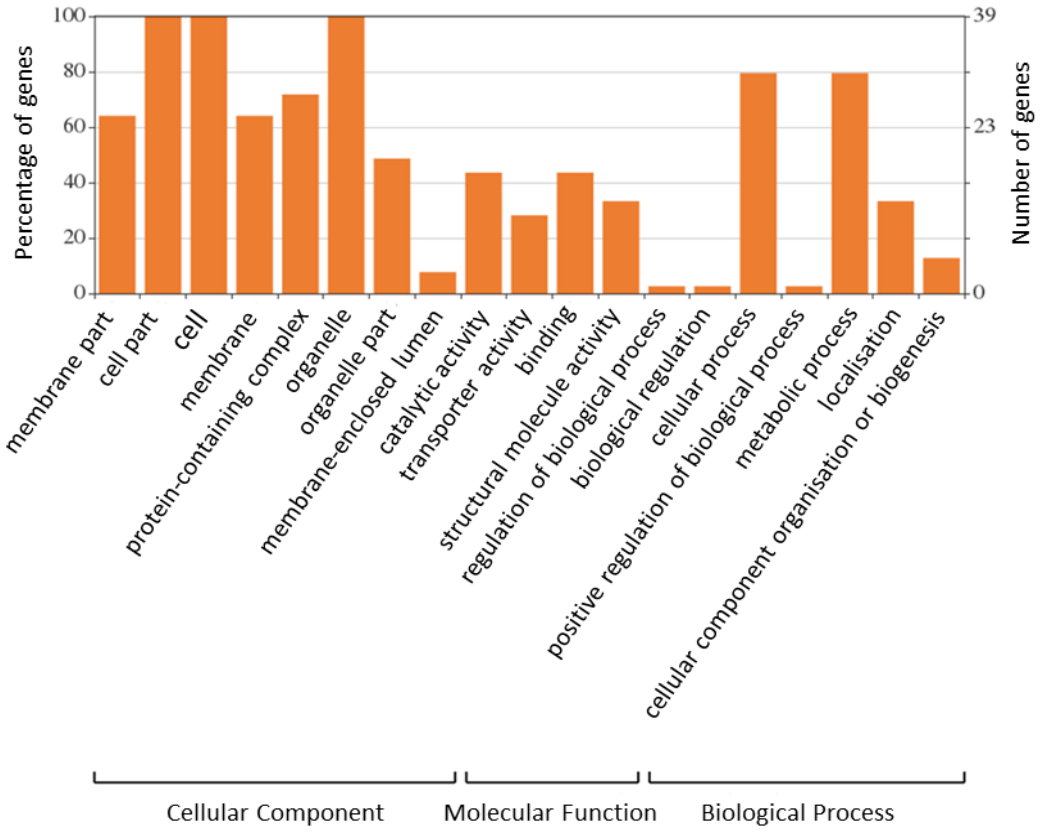

FIGURE 1. Annotation of *T. rafflesiae* mitochondrial genes based on functional groups in three GO categories

COMPARATIVE SEQUENCE ANALYSIS AND PHYLOGENETIC RECONSTRUCTION OF *Tetrastigma* MITOCHONDRIAL GENES

Five mitochondrial protein coding genes (*ccmB*, *cob*, *matR*, *nad6,* and *rps3*) of *T. rafflesiae* were analysed and compared with homologous sequences from three other *Tetrastigma* species. The gene selection was made based on the completeness of the sequences compared to the reference species and the availability of the mitochondrial protein coding gene sequences of the other *Tetrastigma* species in the GenBank database (Table 4). The homologous sequences for each species were analysed separately using Clustal Omega and the percent identity matrix (PIM) for each gene was calculated (Table 5).

Based on the PIM value for each gene aligned, it was evident that a single homologous sequence alignment among *Tetrastigma* species showed very high conservation with no less than 99% identity with a large portion showing 100% identity. These results were expected since coding sequences of plant mitogenomes are characteristically known to exhibit high conservation among species (Alverson et al. 2010; Wynn & Christensen 2019). Subsequently, the five mitochondrial protein coding genes from each species were combined by concatenating the gene sequences and reanalysed with Clustal Omega. Based on the PIM results for the combined sequences (Table 5), a lower conservation was shown with the lowest score of 98.73% and without any 100% identity. This demonstrated that the use of combined mitochondrial gene sequences can increase the nucleotide variant and may be useful in better distinguishing different species of *Tetrastigma*.

TABLE 4. Reference mitochondrial genes of *Tetrastigma* species used in multiple sequence alignment analysis

Gene	Species	GenBank Accession No.
	T. obovatum	EF135178.1
ccmb	*T. rumicispermum*	EF135179.1
	T. voinierianum	EF135180.1
	T. obovatum	EF135286.1
cob	*T. rumicispermum*	EF135287.1
	T. voinierianum	EF135288.1
	T. obovatum	EF135308.1
matR	*T. rumicispermum*	EF135309.1
	T. voinierianum	EF135310.1
	T. obovatum	EF135410.1
nad6	*T. rumicispermum*	EF135411.1
	T. voinierianum	EF135412.1
	T. obovatum	EF135495.1
rps3	*T. rumicispermum*	EF135496.1
	T. voinierianum	EF135497.1

TABLE 5. Percent identity matrix of matches between mitochondrial genes of *Tetrastigma* species

Gene	Species	*T. rafflesiae*	*T. obovatum*	*T. rumicispermum*	*T. voinierianum*
ccmB	*T. rafflesiae*	100	99.81	99.81	99.81
	T. obovatum	99.81	100	100	100
	T. rumicispermum	99.81	100	100	100
	T. voinierianum	99.81	100	100	100
cob	*T. rafflesiae*	100	99.87	100	99.87
	T. obovatum	99.87	100	99.87	100
	T. rumicispermum	100	99.87	100	99.87
	T. voinierianum	99.87	100	99.87	100
matR	*T. rafflesiae*	100	100	100	100
	T. obovatum	100	100	100	100
	T. rumicispermum	100	100	100	100
	T. voinierianum	100	100	100	100
nad6	*T. rafflesiae*	100	100	100	100
	T. obovatum	100	100	100	100
	T. rumicispermum	100	100	100	100
	T. voinierianum	100	100	100	100
rps3	*T. rafflesiae*	100	99.71	100	99.72
	T. obovatum	99.71	100	99.9	100
	T. rumicispermum	100	99.9	100	99.9
	T. voinierianum	99.72	100	99.9	100
Combined	*T. rafflesiae*	100	99.96	99.92	99.93
	T. obovatum	99.96	100	99.87	99.98
	T. rumicispermum	99.92	99.87	100	99.84
	T. voinierianum	99.93	99.98	99.84	100

By using the MSA of the combined homologous mitochondrial gene sequences, we reconstructed a phylogenetic tree using the ML method with 1000 bootstrap replication to identify the phylogenetic relationship among *Tetrastigma* species (Figure 2 & Table 6). This analysis was conducted to distinguish *T. rafflesiae*, which is the host for the parasitic *Rafflesia* from other *Tetrastigma* species. The results showed that *T. obovatum* and *T. voinierianum* are sisters with 89% bootstrap percentage (BP) while *T. rafflesiae* is a sister taxon to the clade with 76% BP. *T. rumicispermum* is positioned out of the node of those three species with 100% BP while *V. vinifera* being a different genus is the outgroup of all *Tetrastigma* species. Overall, the results indicated that *T. rafflesiae* can be differentiated from the other *Tetrastigma* species.

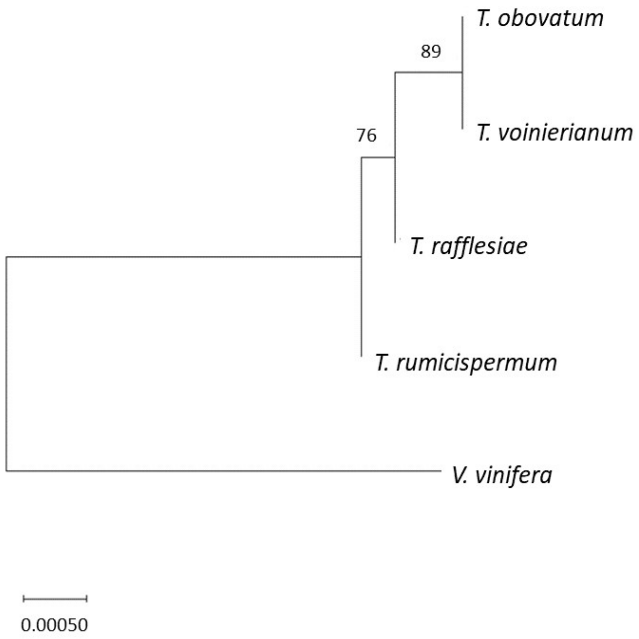

FIGURE 2. Phylogenetic tree reconstruction of *Tetrastigma* species with the outgroup species *V. vinifera.* The combination of five mitochondrial genes of *Tetrastigma* and *V. vinifera* (*ccmB*, *cob*, *matR*, *nad6,* and *rps3*) was used to reconstruct the phylogenetic tree using Maximum Likelihood and GTR model in MEGA X with 1000 bootstrap replications.

TABLE 6. Data matrices of the combined mitochondrial gene sequences used in the phylogenetic tree reconstruction of *Tetrastigma* species

Sequence attributes	
No. of ingroup sequences used	4
Final length after alignment and trimming (bp)	6753
No. of conserved sites	5973
No. of variable sites	104
No. of parsimony-informative sites	13

DISCUSSION

Several *Tetrastigma* species especially those found in China have been reported to confer medicinal benefits based on various phytochemical studies (Fu et al. 2011; Ji et al. 2021; Xu et al. 2008; Yang et al. 1989). Nevertheless, other *Tetrastigma* species like *T. rafflesiae* attracts attention due to its exclusive relationship as a host plant to a rare parasitic plant genus. However, *Tetrastigma* distribution does not overlap only with regions with *Rafflesia* because *Tetrastigma* has a more extensive geographical distribution (Chen et al. 2011; Nais 2001). Yet, the presence of *Tetrastigma* in a shared habitat with *Rafflesia* may be a positive indication that the parasitic plant is also present. This is due to the holoparasitic nature of *Rafflesia*, which depends completely on *Tetrastigma* to survive.

The mitogenome sequence of *T. rafflesiae* was assembled from mitochondrial specific reads identified by mapping whole genome sequencing reads to reference plant mitogenomes. The final length of the assembled scaffold of *T. rafflesiae* mitogenome is 336 kb, which is similar to *A. thaliana* (367 kb) but smaller than *V. vinifera* (773 kb) (Goremykin et al. 2009; Sloan et al. 2018). The gene content for both reference species are similar with *T. rafflesiae*, which comprises of 40 protein coding, 20 tRNA and two rRNA genes. While *A. thaliana* was reported to contain 33 protein coding, 22 tRNA and three rRNA genes, the large mitogenome of *V. vinifera* contains 40 protein coding, 20 tRNA and three rRNA genes based on Motify annotation. This finding showed that despite the difference in mitogenome size, the coding sequences for all three species compared are very similar. This is in agreement with previous studies in which the large variation in the organelle genome size was reported to be mostly attributed to its non-coding DNA content including the repeats, introns, mobile elements and foreign DNA. While the non-coding DNA sequence size varies greatly, the coding sequences seem to be highly conserved (Sloan et al. 2012; Smith & Keeling 2015).

There is a pressing need to preserve the parasitic symbiosis of *Tetrastigma* and *Rafflesia* in nature and the identification of *Tetrastigma* species such as *T. rafflesiae* that could host the parasite would certainly be of value. Previous studies have depended solely on the morphological observation of the host samples collected in the field to determine *Tetrastigma* species (Arshad et al. 2021, 2020; Nasihah et al. 2016; Wan Zakaria et al. 2016). However, molecular approach by the DNA barcoding technique using specific marker sequences could pave for a better alternative to identify different species of *Tetrastigma.* DNA barcoding is a method used for the identification of species based on unique sequences in specific regions of DNA (Hebert et al. 2003). Compared with the conventional morphological identification, the DNA barcoding technique only involves a minute amount of tissue sample for DNA sequencing rather than extensive specimen collections. This is advantageous in lessening the negative impact on the *Tetrastigma* population due to specimen destruction in the field and would consequently support the conservation efforts of *Rafflesia* in its habitat that depends on the survival of the host. Nonetheless, the DNA barcoding technique is only achievable with the accessibility of the genetic sequences of *Tetrastigma* species.

Even though previous plant DNA barcoding studies have preferred chloroplast or nuclear based sequences as markers (Álvarez & Wendel 2003; Chase et al. 2005; Hollingsworth et al. 2009), mitochondrial based markers have started gaining attention due to the advancement of high-throughput DNA sequencing technology that would allow fast reconstruction of mitogenome sequences that would otherwise be challenging to achieve due to their complex structure and variation (Duminil 2014). The mitochondrial based approach has been tested on several plant species including parasitic angiosperms (Barkman et al. 2007). Moreover, comparisons on the application of multigene approach from both mitogenome and plastome sequences to resolve the phylogeny have been observed on basal angiosperm (Qiu et al. 2006), wild octoploid strawberry (Fragaria) species (Govindarajulu et al. 2015) and olive family Oleaceae (Van de Paer et al. 2018). Certain plant genus including *Plantago* has been reported to have an accelerated rate of mitochondrial substitution that exceeds even the fastest animal mitogenomes (Cho et al. 2004), thus making the application of mitochondrial based marker even more promising.

The characterisation of mitochondrial genes of *T. rafflesiae* provides a groundwork for the comparison of several mitochondrial protein coding genes of *T. rafflesiae* with other *Tetrastigma* species. The results of this study demonstrated that compared to individual mitochondrial gene sequences, the multisequence approach by concatenating several mitochondrial genes into one single sequence reduced the identity score and increased sequence variability among different *Tetrastigma* species. This strategy has been observed in previous findings that combined several sequences from mitogenomes for the development of mitochondrial gene based phylogeny of angiosperms (Qiu et al. 2010). Thus, the outcome of this study specifically both the percentage identity values obtained from the comparison of

concatenated mitochondrial genes and the phylogenetic placement of *T. rafflesiae* and other *Tetrastigma* species indicated that the combination of mitochondrial based gene sequences has the potential to discriminate *T. rafflesiae* from other *Tetrastigma* species and can be further developed as species-specific markers.

Conclusions

In this study, the mitochondrial genes of *T. rafflesiae* were identified and characterised. Comparative sequence and phylogenetic analyses of different *Tetrastigma* species showed that combined mitochondrial genes was able to differentiate *T. rafflesiae* from other *Tetrastigma* species studied. The capability to survey and identify the geographical locations of *T. rafflesia* can be used as an indicator for potential existence of *Rafflesia* species and in the future may play an important role for conservation efforts such as relocation of *Rafflesia* populations from threatened regions to a new suitable location with sufficient host presence.

Acknowledgements

This research was funded by UKM Research University Grant (GUP-2016-008).

References

Adam, J.H., Juhari, M.A.A., Mohamed, R., Abdul Wahab, N.A., Arshad, S., Kamaruzaman, M.P., Mohd Raih, M.F. & Wan, K-L. 2016. *Rafflesia tuanku-halimii* (Rafflesiaceae), a new species from Peninsular Malaysia. *Sains Malaysiana* 45(11): 1589-1595.

Aguileta, G., De Vienne, D.M., Ross, O.N., Hood, M.E., Giraud, T., Petit, E. & Gabaldón, T. 2014. High variability of mitochondrial gene order among fungi. *Genome Biology and Evolution* 6(2): 451-465.

Arshad, S., Juhari, M.A.A., Talip, N., Abdul Wahab, N.A.A. & Adam, J. 2021. Anatomy and micromorphology of *Tetrastigma rafflesiae* (Vitaceae). *Sains Malaysiana* 50(4): 897-905.

Arshad, S., Juhari, M.A.A., Talip, N., Abdul Wahab, N.A.A., Fadzilah, S. & Adam, J. 2020. Comparative leaf anatomy of *Tetrastigma rafflesiae* (Miq.) Planchon and *Tetrastigma pedunculare* (Wall. *ex* Laws.) Planch. in Peninsular Malaysia. *Sains Malaysiana* 49(4): 721-729.

Álvarez, I. & Wendel, J.F. 2003. Ribosomal ITS sequences and plant phylogenetic inference. *Molecular Phylogenetics and Evolution* 29(3): 417-434.

Alverson, A.J., Wei, X., Rice, D.W., Stern, D.B., Barry, K. & Palmer, J.D. 2010. Insights into the evolution of mitochondrial genome size from complete sequences of *Citrullus lanatus* and *Cucurbita pepo* (Cucurbitaceae). *Molecular Biology and Evolution* 27(6): 1436-1448.

Barkman, T.J., McNeal, J.R., Lim, S-H., Coat, G., Croom, H.B., Young, N.D. & dePamphilis, C.W. 2007. Mitochondrial DNA suggests at least 11 origins of parasitism in angiosperms and reveals genomic chimerism in parasitic plants. *BMC Evolutionary Biology* 7(1): 248.

Boetzer, M., Henkel, C.V., Jansen, H.J., Butler, D. & Pirovano, W. 2010. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* 27(4): 578-579.

Bosi, E., Donati, B., Galardini, M., Brunetti, S., Sagot, M.-F., Lió, P., Crescenzi, P., Fani, R. & Fondi, M. 2015. MeDuSa: A multi-draft based scaffolder. *Bioinformatics* 31(15): 2443-2451.

Chase, M.W., Salamin, N., Wilkinson, M., Dunwell, J.M., Kesanakurthi, R.P., Haidar, N. & Savolainen, V. 2005. Land plants and DNA barcodes: Short-term and long-term goals. *Philosophical Transactions of the Royal Society B: Biological Sciences* 360(1462): 1889-1895.

Chen, P., Chen, L. & Wen, J. 2011. The first phylogenetic analysis of *Tetrastigma* (Miq.) Planch, the host of Rafflesiaceae. *Taxon* 60(2): 499-512.

Cho, Y., Mower, J.P., Qiu, Y.L. & Palmer, J.D. 2004. Mitochondrial substitution rates are extraordinarily elevated and variable in a genus of flowering plants. *Proceedings of the National Academy of Sciences of the United States of America* 101(51): 17741-17746.

Christin, P.A., Besnard, G., Edwards, E.J. & Salamin, N. 2012. Effect of genetic convergence on phylogenetic inference. *Molecular Phylogenetics and Evolution* 62(3): 921-927.

Corriveau, J.L. & Coleman, A.W. 1988. Rapid screening method to detect potential biparental inheritance of plastid DNA and results for over 200 angiosperm species. *American Journal of Botany* 75(10): 1443-1458.

Donnelly, K., Cottrell, J., Ennos, R.A., Vendramin, G.G., A'Hara, S., King, S., Perry, A., Wachowiak, W. & Cavers, S. 2017. Reconstructing the plant mitochondrial genome for marker discovery: A case study using *Pinus*. *Molecular Ecology Resources* 17(5): 943-954.

Duminil, J. 2014. Mitochondrial genome and plant taxonomy. *Methods in Molecular Biology* 1115: 121-140.

Fauron, C., Allen, J., Clifton, S. & Newton, K. 2004. Plant mitochondrial genomes. In *Molecular Biology and Biotechnology of Plant Organelles*, edited by Daniell, H. & Chase, C. Dordrecht: Springer. pp. 151-177.

Feliner, G.N. & Rosselló, J.A. 2007. Better the devil you know? Guidelines for insightful utilization of nrDNA ITS in species-level evolutionary studies in plants. *Molecular Phylogenetics and Evolution* 44(2): 911-919.

Felsenstein, J. 1981. Evolutionary trees from DNA sequences: A maximum likelihood approach. *Journal of Molecular Evolution* 17(6): 368-376.

Fu, Y-M., Jiang, W-M. & Fu, C-X. 2011. Identification of species within *Tetrastigma* (Miq.) Planch. (Vitaceae) based on DNA barcoding techniques. *Journal of Systematics and Evolution* 49(3): 237-245.

Goremykin, V.V., Salamini, F., Velasco, R. & Viola, R. 2009. Mitochondrial DNA of *Vitis vinifera* and the issue of rampant horizontal gene transfer. *Molecular Biology and Evolution* 26(1): 99-110.

Götz, S., García-Gómez, J.M., Terol, J., Williams, T.D., Nagaraj, S.H., Nueda, M.J., Robles, M., Talón, M., Dopazo, J. & Conesa, A. 2008. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Research* 36(10): 3420-3435.

Govindarajulu, R., Parks, M., Tennessen, J.A., Liston, A. & Ashman, T.A. 2015. Comparison of nuclear, plastid, and mitochondrial phylogenies and the origin of wild octoploid strawberry species. *American Journal of Botany* 102(4): 544-554.

Gualberto, J.M., Mileshina, D., Wallet, C., Niazi, A.K., Weber-Lotfi, F. & Dietrich, A. 2014. The plant mitochondrial genome: Dynamics and maintenance. *Biochimie* 100: 107-120.

Hebert, P.D.N., Cywinska, A., Ball, S.L. & DeWaard, J.R. 2003. Biological identifications through DNA barcodes. *Proceedings of the Royal Society B: Biological Sciences* 270(1512): 313-321.

Hollingsworth, P.M., Forrest, L.L., Spouge, J.L., Hajibabaei, M., Ratnasingham, S., van der Bank, M., Chase, M.W., Cowan, R.S., Erickson, D.L., Fazekas, A.J., Graham, S.W., James, K.E., Kim, K-J., Kress, W.J., Schneider, H., van AlphenStahl, J., Barrett, S.C.H., van den Berg, C., Bogarin, D., Burgess, K.S., Cameron, K.M., Carine, M., Chacòn, J., Clark, A., Clarkson, J.J., Conrad, F., Devey, D.S., Ford, C.S., Hedderson, T.A.J., Hollingsworth, M.L., Husband, B.C., Kelly, L.J., Kesanakurti, P.R., Kim, J.S., Kim, Y-D., Lahaye, R., Lee, H-L., Long, D.G., Madriñán, S., Maurin, O., Meusnier, I., Newmaster, S.G., Park, C-W., Percy, D.M., Petersen, G., Richardson, J.E., Salazar, G.A., Savolainen, V., Seberg, O., Wilkinson, M.J., Yi, D-K. & Little, D.P. 2009. A DNA barcode for land plants. *Proceedings of the National Academy of Sciences of the United States of America* 106(31): 12794-12797.

Ji, T., Ji, W.W., Wang, J., Chen, H.J., Peng, X., Cheng, K.J., Qiu, D. & Yang, W.J. 2021. A comprehensive review on traditional uses, chemical compositions, pharmacology properties and toxicology of *Tetrastigma hemsleyanum*. *Journal of Ethnopharmacology* 264: 113247.

Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. 2018. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution* 35(6): 1547-1549.

Langmead, B. & Salzberg, S.L. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods* 9(4): 357.

Latiff, A. 1983. Studies in Malesian Vitaceae VII. The genus *Tetrastigma* in the Malay Peninsula. *Gardens' Bulletin Singapore* 36(2): 213-228.

Li, X., Yang, Y., Henry, R.J., Rossetto, M., Wang, Y. & Chen, S. 2015. Plant DNA barcoding: From gene to genome. *Biological Reviews* 90(1): 157-166.

Liu, D., Ju, J.H., Xu, X.D., Yang, J.S. & Tu, G.Z. 2002. New C-glycosylflavones from *Tetrastigma hemsleyanum* (Vitaceae). *Acta Botanica Sinica* 44(2): 227-229.

Molina, J., Hazzouri, K.M., Nickrent, D., Geisler, M., Meyer, R.S., Pentony, M.M., Flowers, J.M., Pelser, P., Barcelona, J., Inovejas, S.A., Uy, I., Yuan, W., Wilkins, O., Michel, C-I., Locklear, S., Concepcion, G.P. & Purugganan, M.D. 2014. Possible loss of the chloroplast genome in the parasitic flowering plant *Rafflesia lagascae* (Rafflesiaceae). *Molecular Biology and Evolution* 31(4): 793-803.

Nabholz, B., Glémin, S. & Galtier, N. 2009. The erratic mitochondrial clock: Variations of mutation rate, not population size, affect mtDNA diversity across birds and mammals. *BMC Evolutionary Biology* 9(1): 1-13.

Nais, J. 2001. *Rafflesia of the World*. Sabah: Natural History Publications.

Nasihah, M., Zulhazman, H., Siti Munirah, M.Y., Wan Norqayyum Nadia, W.A. & Latiff, A. 2016. *Tetrastigma hookeri* (Laws.) Planch.(Vitaceae), a host plant for *Rafflesia kerri* Meijer in Peninsular Malaysia. *Malayan Nature Journal* 68: 33-39.

Nikolov, L.A., Tomlinson, P.B., Manickam, S., Endress, P.K., Kramer, E.M. & Davis, C.C. 2014. Holoparasitic Rafflesiaceae possess the most reduced endophytes and yet give rise to the world's largest flowers. *Annals of Botany* 114(2): 233-242.

Paterson, A.H., Freeling, M., Tang, H. & Wang, X. 2010. Insights from the comparison of plant genome sequences. *Annual Review of Plant Biology* 61: 349-372.

Qiu, Y.L., Li, L., Wang, B., Xue, J.Y., Hendry, T.A., Li, R.Q., Brown, J.W., Liu, Y., Hudson, G.T. & Chen, Z.D. 2010. Angiosperm phylogeny inferred from sequences of four mitochondrial genes. *Journal of Systematics and Evolution* 48(6): 391-425.

Qiu, Y.L., Li, L., Hendry, T.A., Li, R., Taylor, D.W., Issa, M.J., Ronen, A.J., Vekaria, M.L. & White, A.M. 2006. Reconstructing the basal angiosperm phylogeny: Evaluating information content of mitochondrial genes. *Taxon* 55(4): 837-856.

Reboud, X. & Zeyl, C. 1994. Organelle inheritance in plants. *Heredity* 72: 132-140.

Sievers, F. & Higgins, D.G. 2014. Clustal Omega. *Current Protocols in Bioinformatics* 48(1): 3-13.

Sloan, D.B., Wu, Z. & Sharbrough, J. 2018. Correction of persistent errors in *Arabidopsis* reference mitochondrial genomes. *The Plant Cell* 30(3): 525-527.

Sloan, D.B., Alverson, A.J., Chuckalovcak, J.P., Wu, M., McCauley, D.E., Palmer, J.D. & Taylor, D.R. 2012. Rapid evolution of enormous, multichromosomal genomes in flowering plant mitochondria with exceptionally high mutation rates. *PLoS Biology* 10(1): e100124.

Smith, D.R. & Keeling, P.J. 2015. Mitochondrial and plastid genome architecture: Reoccurring themes, but significant differences at the extremes. *Proceedings of the National Academy of Sciences of the United States of America* 112(33): 10177-10184.

Sun, Y., Moore, M.J., Zhang, S., Soltis, P.S., Soltis, D.E., Zhao, T., Meng, A., Li, X., Li, J. & Wang, H. 2016. Phylogenomic and structural analyses of 18 complete plastomes across nearly all families of early-diverging eudicots, including an angiosperm-wide analysis of IR gene content evolution. *Molecular Phylogenetics and Evolution* 96: 93-101.

Van de Paer, C., Bouchez, O. & Besnard, G. 2018. Prospects on the evolutionary mitogenomics of plants: A case study on the olive family (Oleaceae). *Molecular Ecology Resources* 18(3): 407-423.

Wan Zakaria, W.N.F., Ahmad Puad, A.S., Geri, C., Zainudin, R. & Latiff, A. 2016. *Tetrastigma diepenhorstii* (Miq.) Latiff (Vitaceae), a new host of *Rafflesia tuan-mudae* Becc. (Rafflesiaceae) in Borneo. *Journal of Botany* 2016: Article ID. 3952323.

Wendel, J.F. & Doyle, J.J. 1998. Phylogenetic incongruence: Window into genome history and molecular evolution. In *Molecular Systematics of Plants II,* edited by Soltis, D.E., Soltis, P.S. & Doyle, J.J. Boston: Springer. pp. 265-296.

Wynn, E.L. & Christensen, A.C. 2019. Repeats of unusual size in plant mitochondrial genomes: Identification, incidence and evolution. *G3: Genes, Genomes, Genetics* 9(2): 549-559.

Xu, C., Ding, G-Q., Fu, J-Y., Meng, J., Zhang, R-H. & Lou, X-M. 2008. Immunoregulatory effects of ethyl-acetate fraction of extracts from *Tetrastigma hemsleyanum* Diels et. Gilg on immune functions of ICR mice. *Biomedical and Environmental Sciences* 21: 325-331.

Yang, X.L., Luo, J., Sun, S.B., Wang, H.Z., Wu, X.Y., Liu, H. & Peng, H.G. 1989. Study on antiviral effect of *Tetrastigma hemsleyanum*. *Hubei Journal of Traditional Chinese Medicine* 4: 40-41.

Yang, Z. & Rannala, B. 2012. Molecular phylogenetics: Principles and practice. *Nature Reviews Genetics* 13(5): 303-314.

Ye, J., Zhang, Y., Cui, H., Liu, J., Wu, Y., Cheng, Y., Xu, H., Huang, X., Li, S., Zhou, A., Zhang, X., Bolund, L., Chen, Q., Wang, J., Yang, H., Fang, L. & Shi, C. 2018. WEGO 2.0: A web tool for analyzing and plotting GO annotations, 2018 update. *Nucleic Acids Research* 46(W1): W71-W75.

Ye, J., Fang, L., Zheng, H., Zhang, Y., Chen, J., Zhang, Z., Wang, J., Li, S., Li, R., Bolund, L. & Wang, J. 2006. WEGO: A web tool for plotting GO annotations. *Nucleic Acids Research* 34(suppl. 2): W293-W297.

Zerbino, D.R. & Birney, E. 2008. Velvet: Algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Research* 18(5): 821-829.

*Corresponding author; email: klwan@ukm.edu.my