IDENTIFICATION OF AN ALBINO SQUIRREL (SCURIDAE) USING DNA BARCODE (COX1) AND PHYLOGENETIC ANALYSIS

by

LEONG JIA SHAN

A thesis submitted in fulfilment of the requirements for the degree of Bachelor of Applied Science (Natural Resources Science) with Honours

FACULTY OF EARTH SCIENCE

UNIVERSITI MALAYSIA KELANTAN

2019

APPROVAL

I hereby declare that I have read this thesis and in our opinion this thesis is sufficient in terms of scope and quality for the award of the degree of Bachelor of Applied Science (Geoscience) with Honors Faculty of Earth Science, Universiti Malaysia Kelantan.

Signature : _____

Name : _____

Date : _____

DECLARATION

I declare that this thesis entitled "Identification of an albino squirrel (Sciuridae) using DNA barcode (COX1) and Phylogenetic Analysis" is the result of my own research except as cited in the references. Final Year Project. The thesis has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.

Signature : _____

Name : _____

Date : _____

ACKNOWLEDGEMENT

I would frankly like to express the deepest appreciation to my Final Year Project supervisor, Dr. Jayaraj Vijaya Kumaran, he who has the attitude and the substance of a genius researcher. He has continually provided his enthusiastic encouragement, his continuous guidance along with his given space of discovery within the safe range. I have deep gratitude towards his intention to let us who work under him to be provided for optimum guidance in order to keep us independent at a certain extent, which allow us to gain experiences and knowledges that favor the career or lab skills we may need in the future without causing any troubles or problems towards the authority of Faculty of Earth Science (FSB) and also the authorities from Universiti Malaysia Kelantan, Jeli campus. Without his guidance and persistent supervising implication, this project would not have been possible.

I would also like to thank my lab-mate who also works and learns under Dr. Jayaraj, Ms Tan Shan Wen. She has provided a lot of support and information sharing that helps me to avoid in encountering the rather problematic laboratory policies that the Faculty of Earth Science amended. She has been a trustworthy lab-mate that Dr. Jayaraj acknowledged and has been a positive reference for me in during my laboratory procedure conducting. She is also a contributive lab partner in all of our discussion in genetic knowledges when we are facing similar problems while carrying out the study.

Apart from that, I would also like to convey appreciation towards all the laboratory assistants that have rendered me their patience and assistance. This is especially conveyed towards Ms. Hasimah Binti Hassan who delicately provided her patience and guidance towards me and other the FYP students. I am very thankful to her that she even stayed during mid-semester breaks to be able to give help and supervision to the FYP students, while other laboratory staffs were absent.

Identification of an Albino Squirrel (Sciuridae) using DNA Barcode (COx1) and Phylogenetic Analysis

ABSTRACT

An albino squirrel with white fur and red eyes is discovered in the forest of Kampung Pauh, Bukit Gantang in Perak. The lack of variety on observation such as fur color and stripes distribution making the species identification based on morphology impossible. Thus, this paper states the result of species identification using the phylogenetic analysis on the albino squirrel's CO1 gene DNA barcoding. The research is carried out in 3 main parts which is the DNA extraction, Polymerase Chain Reaction (PCR) and also phylogenetic analysis. The successive extraction attempt on the the targeted DNA segment amplification through PCR. Then the barcoded sequence of the targeted DNA segment was aligned and used to synthesize a phylogenetic tree. The finding of this study shows that the albino squirrel is genetically close to the species, *Callosciurus prevostii* in a certain extent of similarity which is also known as the Asian tri-coloured squirrel. Finally, the finalized result of this study is able to identify the actual species of the albino squirrel. This study is believed to be able to contribute into more possibilities of genomic studies and molecular biology in future developments and applications.

Pengenalpastian Spesies Tupai Albino (Sciuridae) Menggunakan Barcode DNA (COX1) dan Analisis Filogenetik

ABSTRAK

Seekor tupai albino yang berbulu putih dan mempunyai mata merah telah dijumpai di dalam hutan sekeliling Kampung Pauh, Bukit Gantang di Perak. Disebabkan kekurangan maklumat atas morfologi yang boleh diperhatikan, seperti warna bulu, pengedaran dan corak jalur, pengenalpastian bagi spesies tupai albino tersebut tidak dapat disahkan. Oleh itu, penyelidikan ini menunjukkan hasilan pengenalpastian spesies melalui analisis phylogenetik terhadap barcoding DNA gen COx1 tupai albino tersebut. Penyelidikan ini terbahagi kepada 3 bahagian utama iaitu, pengekstrakan DNA, *Polymerase Chain Reaction* (PCR) dan juga analisis filogenetik. Setelah proses pengekstrakan DNA berjaya, segmen DNA yang disasarkan terus diamplifikasikan. *Barcode* yang dijana daripada urutan nukleotida juga diselaraskan dan digunakan untuk membuat lakaran pokok filogenetik. Penemuan bagi penyelidikan tajuk ini menunjukkan bahawa tupai albino tersebut adalah amat berkaitan dengan spesies *Callosciurus prevostii* atau juga digelar sebagai tupai Asia tiga-warna. Akhirnya, hasil analisasi ini berjaya menunjukkan spesies sebenarnya. Penyelidikan ini diharapkan akan membawa lebih sumbangan kepada peluang perkembangan dan aplikasi penyelidikan genetik.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

µl	Micro litre
COx1	Cytochrome Oxidase Subunit 1
ml	Milli litre
PCR	Polymerase Chain Reaction
dNTP	Deoxynucleotide phosphate
DNA	Deoxynucleic acid
TAE	Tris-acetate-EDTA
EDTA	Ethylenediaminetetraacetic acid
kb	Kilo-base pair
Bp	Base pair
min	Minute
sec	Second
g	Gram
mg	Milli gram
ng	Nano gram
M	Molar
mtDNA	Mitochondrial DNA
CTAB	Cetyl trimethylammonium bromide
EMBL	European Molecular Biology Laboratory
DDBJ	DNA Data Bank of Japan
pmol	Pico-mole

%	Percent
Rpm	Revolution per minute
TE	Tris-EDTA
v	Volts
NaCl	Sodium chloride
W	Watt
°C	Degree celcius

CHAPTER 1

INTRODUCTION

1.1 Background

Albinism is a genetically altered trait that is caused by the disorder of the melanin production of an organism. Animals with this disorder have white fur coat, skin and red eyes due to the absence of color pigments (Carden, Boissy, Schoettker, & Good, 1998). Albinism has been seen as disease to human, while it is considered a type of disadvantage of disability to animals as it causes them to lose their protective coloration against their predator (Searle, 1968).

Around two years ago, a group of villagers which has brought the attention of researchers from Universiti Malaysia Kelantan of the discovery of a white squirrel that is rarely sighted at nearby or any forest in Malaysia. Morphological examination of the sample confirms the specimen to be of genus *Callosiurus*, but the key identification on squirrel, Sciuridae, mainly depends on the fur color and stripes distribution on the belly, side and back of the body.

The characteristics of the subjected squirrel are very similar to the white variant of *Callosciurus finlaysoni* but this species is not found in Malaysia. Other than that, the subject appeared to be having a common symptom of albinism, where the eyes are red unlike the normal eye color of *Callosciurus* (Francis & Barrett, 2008). Considering the factor of albinism and the complications of morphology variation in differentiating squirrel species, taxanomic morphology identification method cannot be used in this case.

Therefore, a genetic method is used instead to identify its generic species. DNA barcoding system will be used to analyze the albino squirrel's relativity in terms of COX1 gene DNA sequences among all the other squirrel species in the form of phylogenetic tree. The results from the analysis will be exhibiting in the form of phylogenetic trees, where species that possess more common characteristics in genetic context will be located nearer to each other. (Woese, 2000).

The study of this research involved extraction of genomic DNA from dissected squirrel body and amplifying the COX1 gene using COX1-specific primer in PCR to generate barcode that allows the comparative analysis towards worldwide database such as Genbank, EMBL, CCDB or DDBJ (Harris, 2003).

The reason for using COX1 gene in this research is because the Cytochrome C Oxidase subunit 1 gene is present in most of the eukaryotes. As mitochondria is generating energy, COX1 gene is frequently transcribed and also well-conserved. Hence, the gene is reported to be feasible in copying from an unknown species for verification purpose (Zhao, Gentekaki., Yi, & Lin, 2013).

1.2 Problem statement

The species of albino squirrel found could not be determine by comparing other species and morphological studies by naked eyes observations. The identification cannot be achieved without comparison with morphology characteristics.

1.3 Objective

The objective of the study is to identify the species of the albino squirrel obtained from Kampung Pauh using phylogenetic analysis of DNA barcoding on COX1.

1.4 Scope of study

The scope of study investigated the taxonomy identification of the albino squirrel using molecular markers and generating a comprehensive phylogenetic tree.

1.5 Significant of study

This research used the DNA barcoding technique to identify the species of albino mammals without relying on taxonomic morphological studies.

CHAPTER 2

LITERATURE REVIEW

2.1 Sciuridae

Squirrels and flying squirrels are both under the family of Sciuridae of Rodentia. Squirrel have always mistaken or thought to be the same as treeshrews, *Tupaia*. However, through a closer look on the significant on morphology differences such as the significant pointed muzzles of tree shrews is one of the key identification feature from squirrels. (Thompson et al, 1997). All of them are sharing a certain degree of similarity in anatomical features including teeth and jaw muscles which is the key identification distinguishing feature. There are plenty of different genera of squirrels under the family of *Sciuridae* that can be found in Malaysian region (Cox & Hautier, 2015). Such as, *Callsciurus prevostii*, *Rubisciurus rubriventer, Sundasciurus brookei*, *Dremomys rufigenis,* and *Ratufa bicolor*. The morphological difference among squirrel species is mainly based on the distinctive fur coloration, patterns and striped on the belly, back or side. The speciation variations on the morphology are also varied according to their distribution on geographic area (Francis & Barret, 2008). This has created a greater complication where genetic analysis from wide geographic on the samples can provide a sturdy information on the speciation of squirrel (Francis & Barret, 2008).

2.2 Albinism

Albinism refers to the symptoms on human or animal morphology caused by mutation of melanin formation process (Carden et al., 1998). The mutation creates difference between albinism and leucism, where albinism is the fully absence of melanin in mostly mammals, but leucism can be described as partial albinism which usually occurs on amphibians and reptiles (Veena, Thomas, Raje, & Durgekar, 2011). The effect of albinism failure upon formation of melanocytes in the skin, hair follicles and also in the eyes (King, 1987). This finally leads to the symptoms of fully white or pink skin and abnormal iris color, mostly red in albino animals. There are two types of albinism in human which is Oculacutaneous albinism (OCA) and Ocular albinism (OA), in which the diagnosed symptoms of OCA will have much more deficiency of melanin than OA. In other words, reduced melanin synthesis system that makes the human patient with OCA susceptible to UV radiation which in turn causing severe sunburn and skin cancers. It also causes defects on human patient's visual system on and causes intolerance to surrounding lights, also known as 'photophobia' (Lee et al., 1994). When albinism occurs on small mammals like Rodentia, the melanin pigment will also be missing in the fur coat, the eyes and also the coats. The cases of albino animals to be found or discover by human is rare due to the intersection of probability regarding the rarity of the mutation occurs naturally and the natural selection by the natural predator (Nedyalkov, Koshev, Raykov, & Bardarov, 2014). This is because the abnormality of the skin or fur color has reduced most of the animal's ability to hide from predators (Searle, 1968).

2.3 Polymerase chain reaction and DNA sequencing

Polymerase chain reaction is a technique that aims to amplify some specific segments of DNA. The reaction needs enzyme such as DNA polymerase to catalyze the DNA replication so that DNA copies can be doubled in every cyclic routine. DNA polymerase needs primer to be attached on its 5' side. Primers functions as an activator for the enzyme and also controlling the amplified the targeted region. There are always primers needed to be applied on both side of the strands, because DNA replication has to occur on both strands each is complementary to another and it is able to encircle the region is aimed to be amplified. (Laake, Benestad, & Olsen, 2007). PCR is usually carried out by thermocycler that shift rapidly between temperatures. The thermocycler can regulate the temperature, lengths of each step and number of cycles by the automated program. The first two cycles double the targeted DNA each cycle, and in 30 cycles the number of DNA copies increases exponentially.

DNA sequencing is an analyzing method towards DNA sequences or molecules including PCR products. Analyzing on reaction mixture of PCR as studied by Sanger uses dideoxynucleotides that incorporates into newly produced DNA strands which terminates the further elongation and also activated other termination sites. Single-stranded DNA fragments with different lengths in a discrete spectrum will also be generated (Eid et al., 2009).

While, the order of the chain lengths will be exhibited after electrophoresis the order bases in the template DNA will also be shown. In this phase, the four di-deoxynucleotide phosphate (ddNTP) labelled with fluorescent marker will be separated by electrophoresis in narrow tubes when it is in full spectrum. The laser source and detectors of each color is the method to identify the presence of different fluorochromes that corresponds to their fragment lengths. These combinations will be directed into the computer where the pattern of DNA sequence will be translated into nucleotide information such pairwise distance that needed to synthesize DNA barcodes (Thompson, Seligmann, & Gordon, 2014).

2.4 DNA Barcoding

The concept of DNA barcoding was introduced in an international scientific conference in February 2005 by a group of scientists led by Paul Herbert (Savolainen, Cowan, Vogler, Roderick, & Lane, 2005). This method was very controversial concept when it was introduced internationally that using short DNA sequences as a universal locus for interspecies or intraspecies identification and comparison (Austerlitz et al., 2009). Barcoding techniques provides accessible pathway for professionals and researches towards the biological attributes of any organisms on the earth. The DNA barcoding technique requires a comprehensive genetic database in the form of DNA barcode. In order to create a reliable and accurate database, countless of DNA barcoding needed to be taken into record until the accumulation constructs a database that can describe a whole diverse taxonomy of different kingdoms by parts into it.(Meyer & Paulay, 2005). DNA barcoding provides an efficient

method to identify the genetic variation at species level by implementing the fact that genetic variation between species exceeds within species (Hebert, Ratnasingham, & de Waard, 2003).

The development of DNA barcoding has greatly improved the average performance for the process identifying species. Comparing to the taxonomic identification approach that involves much more complicated work which is also expensive and more time-consuming, (Valentini, Pompanon, & Taberlet, 2009) also stated that this technique will also enable the species composition of environmental samples (Bohmann et al., 2014). Furthermore, as the development kept carried out to optimize the DNA barcoding technique, the Consortium for Barcode of Life (CBOL) has provide a specific database for DNA barcoding, to overcome the trouble commonly occurs in other DNA barcoding reservoir such as Genbank, EMBL or DDBJ including sequencing errors, contaminations, sample misidentifications or taxonomic problems (Kress, Wurdack, Zimmer, Weigt, & Janzen, 2005).

2.4.1 COX1 Gene

Recent DNA barcoding research work of animals or certain fungi prefer mitochondrial genes rather than nuclear gene. COX 1 gene is one of the evolutionary rate locus that can be used gene marker for DNA barcoding from mitochondrial DNA. Mitochondrial DNA was chosen more often rather than nuclear DNA, due to the mtDNA's high evolution rate, which also have high exposure on genetic recombination and its high copy number (Luo et al., 2011). The ideal barcoding gene that is extensively used by the genetic or taxonomic study on eukaryotes including mammals, COX1 gene (Cytochrome C Oxidase subunit 1), which is also from a mtDNA. The COX1 gene is present in most of the eukaryotes. As mitochondria is generating ATP, COX1 gene is always transcribed and also well-conserved. Hence, making the gene to be feasible in copying from an unknown species for verification purpose (Zhao et al, 2013). The COX1 gene possess characteristics of diverse range of functional domain and their high copy numbers per cell that allows standard barcoding procedure to occur on most of the taxa and hence, constructing a comprehensive set of archival data (Smith, Poyarkov, & Hebert, 2008).

The COX1 gene barcoding is contributing to the achievement of DNA barcoding as an internationally standardized tool for large variation between species identification technology that requires only some standardized short sequence (Austerlitz et al., 2009). By applying the COX1 gene into the barcoding technique, that mainly depends on the core barcode region for animals. It can generate the species identification resolution level up to 95% or more consistently. Furthermore, it is applicable to wider taxa. These nucleotide bases are of course the yield from the amplification through the process of Polymerase Chain Reaction (PCR) upon the mitochondrion gene extracted from liver cells or white

muscles cells. This is how the standard phylogenetic information are prepared, and the accumulation of various other barcodes accordingly to their taxonomy, allowing a more genetic-perspective outlook towards the organism speciation. Which in other words, the application of DNA barcoding enables the construction of a reference basis-data upon the selected group of organisms and finding the subject's species it belongs to through the phylogenetic relationship (Sivakumar et al., 2013) which will be shown through the analysis.

COX1 gene codes the generation of protein involving in electron transport chain, a mechanism included in oxidative phosphorylation (Dennerlein & Rehling, 2015). COX1 gene has the properties that its biomolecular structure is surrounded by highly conserved primer site which has been discovered to be widely susceptible to various targeting PCR (Luo et al., 2011). This allows up to 650 bp DNA fragments of the COX1 gene, also known as the "Folmer Region" to be amplified easily and also in abundance synthesizing a smooth process of sequencing with resolution up to 95% (Seifert et al., 2007). The proximity of the species-level resolution can be proven in a research carried out by (Hebert et al., 2003) where the research of evaluation on divergence of COX1 among congeneric groups of animals.

2.5 Phylogenetic Analysis

2.5.1 Neighbour-Joining

Phylogenetic analysis commonly presented by the forming of phylogenetic tree, where Neighbour-Joining method is one of the phylogenetic analysis that reconstruct phylogenetic trees that simulate the original phylogenetic tree known or recorded in previous taxonomic researches (Saitou & Nei, 1987). Neighbour-Joining is used in constructing the phylogenetic tree that represents the minimum evolution topology according to lineage in the most simple method of accounting the similar characters with the ancestor species (Saitou & Nei, 1987). By inputting the distance matrix calculated from the distance of DNA or protein sequence difference from the basal species, the algorithm will be used to form branches and relocating each species node position in the tree (Delsuc, Brinkmann, & Philippe, 2005). The algorithm can be calculated in the software of MEGA 7 and other phylogenetic analysing softwares as it is a very common used basic analysing method.

2.5.2 Maximum Parsimony

Maximum parsimony analysis promotes the selection of the simplest possible phylogenetic tree that optimally explains the data (Farris, 1970). The input of this analysis is the trait or character shared by two or more groups of taxa diverged from a common ancestral form and without presuming the common morpho-characteristic which in other words, homoplasy as the connection between species which is a typical characteristic of cladistic pathway (Campbell, & Reece, 2008) of analysis but it also mainly exhibits the characteristic of phenetic which maximum parsimony analysis is also possessing a

transformed distance approach and the neighbors relation method (Delsuc et al., 2005). The calculation of maximum parsimony can only be accounted with the evidence provided on synamorphies and it propagates that the elements of rare homoplasies should be filtered-off for the sake of wide variety of application on phylogenetic analysis (Sober, 1983). The tree of Parsimony is considered non-parametric because the tree generated itself is evaluated on the basis of general metric only (Arnaud & Bamber, 1988). However, the Parsimony method has a major weakness where the maximum parsimony analysis suggests multiple phylogenetic trees are alike and the algorithm does not recommend the best tree out of all the alternatives without the aid of bioinformatic soft wares that tabulates the probability and repetitive (Gaut & Lewis, 1995).

2.5.3 Maximum Likelihood

Maximum Likelihood is a phenetic analyzing method that is well-developed in a statistical study. Maximum Likelihood analysis is the most general estimation method. In a given set of finite parameter, it computes and quantifies the probability and extend of the likelihood in sample data by making an assumption on the population distribution (Enders, 2005). Maximum Likelihood estimates the computation of the likeliness, in characters or significants taken account in the population and also the population estimated mean and variance through the Maximum Likelihood Estimation system, where this has been used to initiate a Maximum Likelihood analysis method to an unbiased population (Freckleton, Harvey, & Pagel, 2002). The neighbor joining implementation is in a restricted manner, the situation needs to be provided with the values of mean and variance estimated from population, which is the algorithm parameters that determine the significant phylogenetic dependence exhibited from the phylogenetic tree (Freckleton, Harvey, & Pagel, 2002). From the aspect of phylogenetic tree analysis, maximum likelihood can also out-perform other methods even given short DNA sequences and the tree topologies of maximum likelihood can also be easily evaluated. However, there are downside of this method, which is the limit of population input and the time consumption of the calculation through its algorithm and this will expectedly be causing intensive stresses to the CPU when running in a software.

CHAPTER 3

MATERIAL AND METHODOLOGY

3.1 Material & apparatus

Apparatus:

The equipment used in this research project were Eppendorf mini-Spin plus centrifuge, Laboff fume-hood, Memmert Water bath, BIO-RAD Power Pac Basic electrophoresis machine, eppendorf micropippete in 10µl, 100µl and 1000µl, Samsung Microwave oven, Thermocycler, beaker, conical flask, reagent bottles, falcon tubes, microcentrifuge tube, PCR tubes, micropippete tips, Sartorius semi micro & analytical balance, Smith vortex mixer MX-S, Protein Simple Alpha Imager HP UV transilluminator, Kyratec Super Cycler Thermal Cycler, Golden Cross dissection kit

Material:

The material used in this study were pre-mixed CTAB solution, 20mg/ml Proteinase K, Chloroform, Isoamyl alcohol, 70% ethanol, 5M NaCl solution, deionized distilled water, 10% TE buffer, Go Flexi-Taq PCR buffer, 1X TAE buffer, ProMEGA 1kb DNA ladder, sodium acetate, absolute ethanol molecular grade (100%), dNTP, *Taq* DNA polymerase, $MgCl_2$, Primer (10 M), LCO1490, HCO2198, Agarose powder, Redsafe nucleic acid staining solution, , albino squirrel tissue sample.

3.2 Cell isolation

The sampled albino squirrel has been dissected with a sterilized scalpel, a small slice of the liver is cut out by scissor and inserted into each microcentrifuge tube and macerated in 70% ethanol. The liver tissue collection has been done in a sterile condition and preserved.

3.3 DNA Extraction & Isolation

After the liver tissue in the microcentrifuge tubes have unfrozen, then each tissue sample were suspended in 250 µl of 5M NaCl solution for more than 5 minutes. Then each tube was added with 250 µL of CTAB isolation buffer pre-warmed up to 60°C added with Proteinase K. After that, the tissue samples were grinded along with the solution and stirred in the tube. Then, all tube with the sample solution have been incubated at 60°C for 30 minutes. After taking the samples out from the incubator, the samples were allowed to cool off and added with 10 µL of 10% SDS solution which followed by incubation at 60°C for 10 minutes. Each tube of the samples was then added with 100 µL of solution mix of chloroform: isoamyl alcohol (24:1) into the sample and mixed well. The samples were then centrifuged at maximum speed, 14,500 rpm for 10 minutes (Ren & Miao, 2018). After centrifugation, the aqueous phase on the top from each tube were removed and relocated it into the new set centrifuge tubes. Afterward, adequate amount with CTAB solution, 250 µL of absolute ethanol was added with 50 µL of the 5M NaCl solution to the aqueous phase of each centrifuge tube and centrifuged. The tubes were then kept in a chiller at -20°C. After the tubes were taken out from chiller, the unfrozen samples were brought to centrifuged again at maximum speed for 10 minutes. The alcohol and salt solution were poured out carefully and replaced by 70% ethanol to wash the excessive salt.

Finally, The DNA pellets in each tube were resuspended in 50-60 µL of TE buffer to saturate the DNA pellets, and each sample of DNA were taken in the volume of approximately 12.5 µL to undergo PCR (Kumaran, 2009). Before the DNA samples undergo PCR, the samples were examined for their DNA concentration using DeNovix, DS-11+ Spectophotometer and recorded.

3.4 Gel electrophoresis

Approximately 0.5 g of agarose powder are weighed and mixed with 50ml of 1x TAE buffer and 5µL of Redsafe nucleic acid staining solution to make up 1% agarose solution, the solution is then heated in microwave oven at 450W with each 30 seconds interval after weighing the total mass before heating (Sabris, 2010). The agarose solution was taken out and observed while swirled every time an interval ended to prevent boiling. The heating process stopped when the agarose solution showed no cloudy sediments and the agarose powder was fully dissolved in the solution. The cleared solution was then weighed at the electronic balance to estimate the weight lost during the heating process. The weight lost after heating was refilled back with approximate amount of 1x TAE buffer. Then the agarose solution was poured into the mould and allowed to be cool down for more than 10 minutes to solidify.

The solidified agar was then placed into the electrophoresis plastic tray filled with 1x TAE buffer. The 2.5 µL DNA ladder and 5 µL of each DNA samples were mixed well with 1 µL of loading dye respectively. Then each sample with loading dye were pipetted into the wells of the agar followed after the ladder. The gel containing dyed DNA samples and 1kb DNA ladder were allowed to run electrophoresis with 90v for 45 minutes by using the BIO-RAD PowerPac Basic electrophoresis machine. The gel was then observed under UV light radiator after electrophoresis was completed to attain the presence of DNA bands. The results were also photographed and recorded (Haines, Tobe, Kobus & Linacre, 2015)

3.5 Polymerase Chain Reaction PCR

The Polymerase Chain Reaction has been carried out with the mixes of reagents of 15.5 µL of ddH_20, 25 µL of PCR buffer, 0.2mM deoxynucleoside triphosphate （dNTP）, 1.5mM $MgCl_2$, 10µM of each CO1 gene primers, LCO1490 (5'GGTCAACAAATCATAAAGATATTGG-3') and HCO2198 (5'TAAACTTCAGGGTGACCAAAAAATCA-3') (Park, Suh, Oh, & Hebert, 2010). Along with and 6 µL of minimum 100 ng/ µL of template DNA and 10x *Taq* polymerase accordingly. The PCR cycle was carried out at firstly on 95°C for 10 minutes in order for the DNA to carry out initial denaturation. After the main denaturation cycle at 95°C for 45 seconds, it was followed by the annealing process for 90 seconds at 50-60 °C after optimizing the specific reaction temperature and the elongation process was carried out at 72°C for 90 seconds. This cycle has repeated for 30 times and ended with a final elongation process for 7 minutes at 72°C respectively (Kumaran, 2009).

Then the visualization of amplified products are observed along with 1 kb DNA ladder on 1% agarose gel containing Redsafe nucleic acid staining solution and allowed to run at 90v for approximately 45 minutes. Then the result on the agarose gel radiated from UV transilluminator was recorded. The amplified sequences were sent to private bio-analyzing corporation, Apical Scientific Sdn Bhd to obtain DNA sequencing data.

Table 3.1 shows the PCR reaction mixture's proportion and their respective concentration and volume in a 25 µL sample.

PCR reagents (Stock Concentration)	Amount for one reaction	
	Volume (µL)	Concentration
ddH_2O	10	
Go Flexi-Taq PCR buffer	2.5	1X
dNTPs (2.5mM)	2	0.25mM
Primer (10 M)		
LCO1490	1	1pmol/ µL
HCO2198	1	
Taq DNA polymerase (5U/ µL)	1	1U/ 25µL
$MgCl_2$ (25mM)	1.5	2mM
Genomic DNA (0.1µg/ µL)	6	>100 ng/ µL
Total	25	

Table 3.2 shows the table of temperature distribution of the 3-step PCR profile.

PCR Processes	Temperature (°C)	Duration	Number of cycles
Initial Denaturation	95	10 min	1
Denaturation	95	30 sec	
Annealing	51.4	30 sec	30
Extension	72	30 sec	
Final Extension	72	5 min	1
*Holding in ice	4	5 min	

3.6 DNA sequence analysis

The sequence obtained from sequencing center has been used to BLAST and \ obtained the top BLAST hits on subject sequences from Genbank in order to carry out nucleotide alignment. The sequencing results of this research were displayed in form of chromatogram files by using the MEGA software in version 7 (Kumar et al., 2016). The MEGA software has also been used to carry out multiple sequence alignments of all DNA sequences of the species shortlisted in matches of BLAST based on B_1's nucleotide sequence using Clustal W (Thompson et al., 1997) including, *Prosciurillus weberi*, *Rubisciurus rubiventer, Sundasciurus lowii*, *Callosciurus prevostii*, *Ratufa bicolor* and etc as shown in **Table 3.3**. Then the analysis has proceeded to the utilization of Kimura-2 parameter (Kimura, 1980) in generating Neighbour-Joining phylogenetic tree (Saitou & Nei, 1987) that represents the model of evolution formed form pairwise genetic distance with 10,000 bootstrap replications (Felsenstein, 1985). The observation of the phylogenetic tree is optimized and rooted using the rodent benchmarking species, *Rattus rattus.* The result in species identification of the albino squirrels was obtained from the phylogenetic tree's observation rooted with the sequence of *Rattus rattus*.

Table 3.3 shows the list of the top subject sequences of BLAST hits species with their respective ascension number from Genbank.

Subject sequences of BLAST hits species	Ascension number from Genbank
Sundasciurus brookei	KP120729, KP994921, ARH02624
Callosciurus erythraeus	HM031932, HM031934, AEB66641
Urocitellus richardsonii	JF457124, AKL82667, NC031209
Callosciurus prevostii	JF459623, JF444288, NC035816
Callosciurus finlaysonii	LN899426, LN899427, LN899428
Callosciurus notatus	KY117541, HM102291
Dremomys rufigenis	KX171264, NC026442, KC447304
Hylopetes alboniger	KX710106, NC031847, JQ601651
Prosciurillus weberi	ALP86225, ALP86238, KR911789
Rattus rattus	HM217729, HM217730, FJ355927
Ratufa bicolor	KF575124, NC023780, KY117582
Rubisciurus rubriventer	KR911790, KR911791, KR911792
Sundasciurus lowii	JF444470, JF459872, JF459873
Urocitellus richardsonii	KP698976, NC031209, JF457124

CHAPTER 4

RESULTS & DISCUSSION

In order to get the sufficient concentration of DNA to carry out Polymerase Chain Reaction (PCR), various attempts of optimization on the CTAB extraction method has been taken to overcome the issue of low DNA concentration and unsuccessful screening when using gel electrophoresis qualification.

4.1 CTAB DNA extraction

The cell membrane and the organelle membranes were broken down and organelles were released out when CTAB reacted to the phospholipid layer. While, SDS solution were also added after and incubated for 5 minutes along with the broken-down tissue samples to mainly inhibit the enzyme, Proteinase K and histone from degrading the DNA of the sample by binding to them. SDS also serves the purpose of denaturing protein and DNA and providing the denatured protein with net negative charge. When the SDS reacts with the salts added along with the alcohol after the incubation, it facilitates the precipitation of DNA which makes it insoluble in the alcohol (Rueckert & Dunker, 1969). According to Tripathy et al (2017), the genomic DNA yield is higher than other optimized method.

4.1.1 Chloroform and isoamyl separation

By adding mixture of chloroform and isoamyl alcohol in the ratio of 24:1 of the volume of CTAB solution added into the microcentrifuge tube that contained the reacted tissue sample. It separated the sample into aqueous phase and organic phase. The proteins and undigested lipid and organelles were in the organic phase because it does not dissolve in water like DNA does in aqueous phase (Wassarman & Kornberg, 1990). While, iso-amyl in the portion of 1/ 25 will function as a detergent that mainly helps reducing foaming while aqueous phase is separated from the organic phase. This results in two layers that can be distinguished by naked eyes observed in the microcentrifuge tube (Tripathy et al., 2017).

4.1.2 Ethanol Precipitation

The alcohol is added to the aqueous phase after centrifugation in the ratio of 1:1 and sodium acetate or sodium chloride are usually chosen to carry out the precipitation where it binds to the DNA make it insoluble in the alcohol. By neutralizing the charges on the nucleic acid molecular structure to the isoelectric point, the salt also serves the purpose in increasing the mass of the DNA-protein complex molecules that helps the progress of precipitation. Sodium acetate or sodium chloride is usually a good choice of salts in precipitating DNA because of its high ionic strength (Kirby, 1956).

However, the process of electrostatic attraction mentioned above is not possible to carry out in a situation where water is the solvent. This is because alcohol have a much lower dielectric constant than water that makes the interaction between Na^+ from salt and PO_3^- from nucleic acid molecule easier to carry out (Nozaki & Tanford, 1971). The

choices between ethanol and isopropanol is controversial, because both of them favored different situation. If ethanol is used in precipitation, the incubation in cold, -20°C is necessary and it takes longer time to precipitate the DNA molecules (Davis et al., 1980) However, it is much easier to get rid of when leave for air drying in room temperature due to its volatile properties. On the other hand, isopropanol, or iso propyl alcohol has higher insolubility for DNA due to its properties that allows it to react with more water molecules. DNA is fully soluble in water molecule due to the charges in water molecules that allows to react with the respective charges on nucleic acid. Isopropanol reacts with more water molecules because of its molecular structure allow it to have dipole-dipole bonding, hydrogen bonding and London dispersion (Mattos & Ringe, 2001). These properties make DNA more insoluble in isopropanol compare to ethanol. Isopropanol then need shorter time to precipitate and does not need to incubate in cold temperature (Cenis, 1992). This comes with the disadvantage of higher evaporating point. It is harder to air dry the isopropanol at the end of the precipitation stage. After comparing two types of alcohol in precipitation stage, the optimized decision is using isopropanol, because it is better to have shorter process in order to ease the other steps in this research and there will be another washing step for the DNA pellets, which can further guarantee the purity of DNA sample which will be later visualized using the 1% agarose gel electrophoresis.

In **Figure 4.1**, the gel electrophoresis result obtained from the UV transilluminator shows the result of three DNA samples extracted from the albino squirrel liver tissue using CTAB extraction method. The three extracted DNA samples, JS 1, JS 2 and JS 3 respectively from left to right next to the 1kb DNA ladder in the result of gel electrophoresis using 1.0% agarose gel and stained with Redsafe nucleic acid stain is as shown in **Figure 4.1**. The result of the gel electrophoresis shows 2 out of 3 DNA samples showed prominent DNA bands down the lane of the inserted the DNA sample, which is JS1 and JS 3. JS 3 is visibly more prominent than JS 1 under the illumination of UV rays.

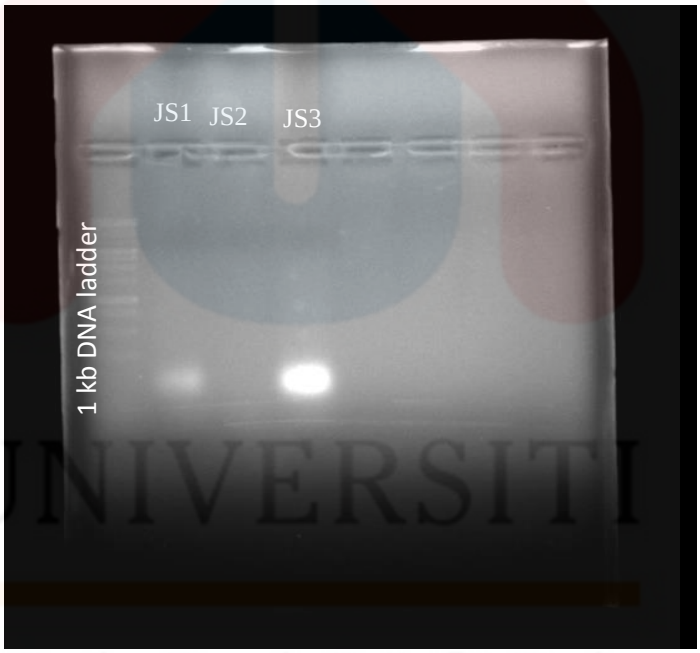


Figure 4.1 shows the gel electrophoresis made by 1.0% agarose gel and stained with Redsafe nucleic acid stain result of DNA samples extracted using CTAB method. Each well contained 6µL of genomic DNA samples. 2 DNA samples extracted JS1 and JS 3 shown DNA bands in the electrophoresis

The concentration of extracted DNA samples JS 1, JS 2 and, JS 3 were quantified using the Nanodrop spectrophotometer accordingly. The readings of the quantification test are listed in the **Table 4.1**. The purity of the nucleic acid content extracted is also shown in the table from the reading of ratio of absorbance at 260:280 nm.

Table 4.1 shows the reading of quantification test of the extracted DNA samples using the DeNovix, DS-11+ Spectophotometer

Sample	Nucleic Acid Concentration (ng/µL)	Factor	A260	260/280	260/230
JS 1	180.530	50	3.6106	1.05	1.76
JS 2	59.246	50	1.1849	0.40	1.30
JS 3	689.647	50	13.7929	1.67	1.33

The extracted DNA sample that has concentration above 100 ng/µL will be expected to show DNA bands from gel electrophoresis after PCR carried out. The absolute purity of the nucleic acid from samples according to the absorbance level from 260:280 nm (Glasel, 1995). It showed that the purity of the DNA samples can be sorted by the descending purity level in the sequence of JS3, JS1 and JS2. Which means the sample with highest purity of nucleic acid is JS 3.

4.2 Polymerase Chain Reaction

The Polymerase Chain Reaction is a method of *in vitro* amplifying a specific gene by synthesizing millions of copies of the targeted segment of DNA. Amplification of DNA segments makes the detection of pathogenic virus or bacteria, identification of individuals and other DNA manipulation research possible (Faloona & Mullis, 1987). By utilizing the thermocycling machine, the number of cycles including denaturation, annealing, extension on DNA can be adjusted as much as needed (James et al., 1994). The PCR is carried out in this study to amplify the COX1 region of the mtDNA from the albino squirrel, using the Universal COX1 barcoding primers, LCO 1490 and HCO 2198 (Vrijenhoek, 1994). The successful amplification of the COX1 region can be determined from the visualization of the DNA bands through 1% agarose gel electrophoresis which halts at the position around 650 bp in given circumstances of 90W current carried throughout the 40 minutes of gel electrophoresis. This is because the complete COX1 gene region comprises up to 658bp (Van Houdt et al., 2010). The optimum temperature for the primers to carry out annealing and also the result and method to determine the integrity of PCR product through gel electrophoresis will be discussed in the following discussion.

4.2.1 Optimizing annealing temperature

Annealing phase is the essential step for PCR, because it affects directly the effective annealing rate of the primer. In this case, the annealing temperature of the primers LCO 1490 and HCO 2198 needs to be optimized, so that the amplification of COX1 gene DNA is at its highest efficiency.

In order to find the optimum temperature for the annealing, a gradient test of PCR reaction was carried out at the temperature from 50-60°C using the template DNA from both JS 1 and JS 2 in two gradient PCR reactions. The gradient temperature range was determined from the estimation accordingly to the fact that the optimum annealing temperature is around the 5 °C below the melting point of the primers (Roux, 2009). From the 12 wells on the gel electrophoresis, only one of it; the third well to be specific has shown a prominent band under the visualization of 1% agarose under UV illuminator. Both gel electrophoresis visualization on PCR product of JS 1 and JS 2 is also highly similar. According to the gradient temperature distribution amongst the 12 wells given, the DNA bands on the third lane falls under the temperature range of 50.9 - 51.8°C. It is safer to pick the medium point of the temperature range, which is 51.4°C to begin with in order to examine the exact optimum annealing temperature for the primer.

Based on the result of visualization on the PCR products in the following PCR process, the quality of DNA bands and its prominent observation is a convincing proof that the temperature point, 51.4°C would be the inferred as the optimum annealing temperature verified for the primer, LCO 1490 and HCO 2198. Although, the formation of DNA bands in an observation in agarose gel visualization does not rely only on the annealing temperature, but it is convincing enough to be determined as a constant variable

in order to regulate the PCR reaction mixture's proportion and adjust all the other uncertain variables. Hence, the optimum annealing temperature of 51.4°C is verified for the primer, LCO 1490 and HCO 2198.

In order to find the optimal annealing temperature for the primer LCO 1490 and HCO 2198 to carry out effective PCR process. The gradient temperature test of PCR is carried out on both DNA samples, JS 1 and JS 3 and the PCR product is visualized on the 1% agarose gel. The result of PCR gradient test for JS1 and JS 3 are shown respectively in **Figure 4.2** and **4.3**.

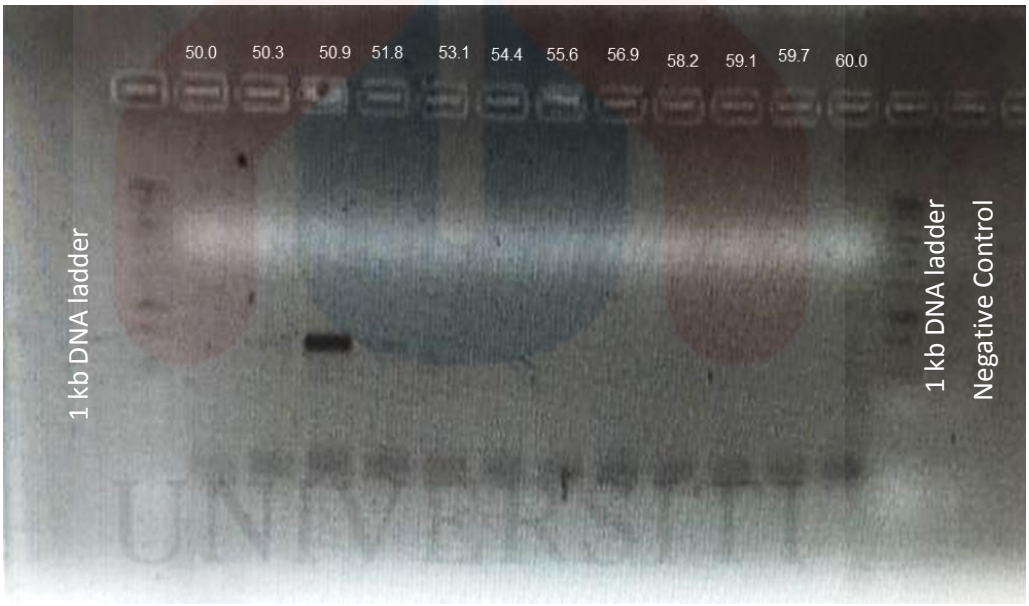


Figure 4.2 shows the visualization of PCR product of DNA sample in a 1% agarose gel electrophoresis, the genomic DNA sample JS 1 in the gradient test over the temperature of 50°-60°C. 3rd lane labelled with 50.9°C is chosen as the optimal annealing temperature. 1kb DNA ladder is contained on the 1st lane and the 13th lane and the negative control is located beside the 13th lane's 1kb DNA ladder.

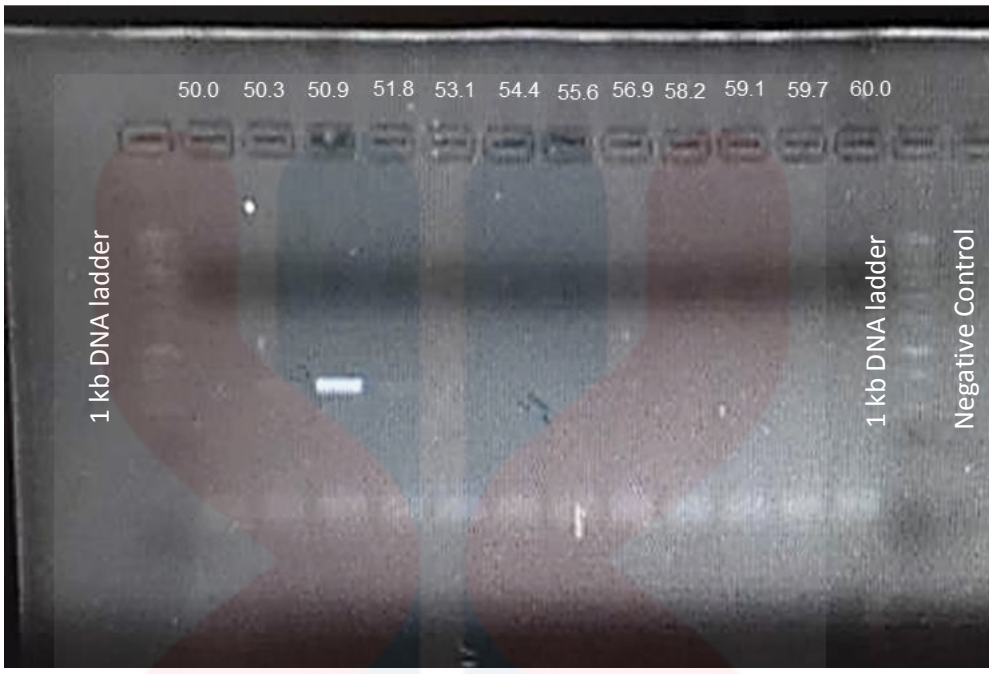


Figure 4.3 shows the visualization of PCR product of DNA sample in a 1% agarose gel electrophoresis, the genomic DNA sample JS 3 in the gradient test over the temperature of 50°-60°C. 3rd lane labelled with 50.9°C is chosen as the optimal annealing temperature. 1kb DNA ladder is contained on the 1st lane and the 13th lane and the negative control is located beside the 13th lane's 1kb DNA ladder.

Both the result from gel electrophoresis visualization shows the third well out of 12 of them exhibited a significant DNA band. This means that the gradient temperature distribution across the 12 well from 50°- 60°C applied on the annealing temperature have coincide one temperature level that is optimum for the primers, LCO 1490 and HCO 2198 to anneal and elongate at its highest efficiency comparatively to other temperature level in the gradient. This is directly related to the gradient temperature distribution of the 12 wells in a column used in thermocycling machine. The temperature distribution applied on the PCR samples in PCR tubes across the gradient 50°- 60°.

The optimal temperature of annealing for primers LCO 1490 and HCO 2198 in PCR carried out on the DNA sample JS 1 and JS 3 falls within the range of 50.9°C and 51.8°C. Therefore, the middle point of the gradient temperature, 51.4°C is taken as the standard annealing temperature applying on the PCR carried out in next step respective to the LCO 1490 and HCO 2198. The amplification of the targeted DNA segments needed to have various elements involved in order to get significant and good quality DNA bands to be shown in the visualization on agarose gel under the UV illumination. In this study, the PCR profile is needed to be arranged so that primer dimer situation would not hinder the effective PCR occurs on the targeted DNA segments and also avoid unsuccessful PCR results. The optimum annealing temperature respective to the primers used needs to be determined through gradient temperature test on the annealing temperature during PCR process. The PCR that has been carried out after the gradient test was using the annealing at the temperature of 51.4°C as discussed. Each tube of PCR reaction mixture sample prepared is 12.5µL in volume, and each pair of PCR reaction mixture sample in PCR tube is transferred from a DNA sample extracted before. The samples that are subjected to the current PCR process is A_2 and B_2, where the template DNA is taken from DNA sample JS 2, along with A_1 and B_1, A_3 and B_3 where the DNA templates are taken respectively from JS 1 and JS 3. The PCR occurred under the thermocycling profile as shown in **Table 3.2**. Similarly, the amplified DNA of the PCR product is once again visualized on the 1% agarose gel and the result is shown in **Figure 4.5** below with the arrangement of samples in the order of A_2, B_2, A_1, B_1, A_3 and B_3 accordingly.

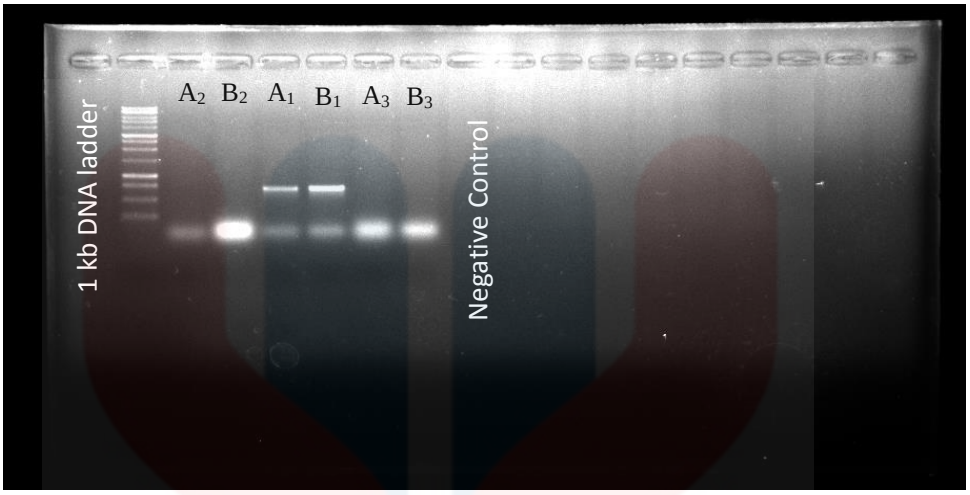


Figure 4.4 shows the result of 1% agarose gel electrophoresis visualization of the 12.5 µL PCR product. Two prominent bands can be observed on the 5th lane and 6th lane from left, 1kb DNA ladder is contained on the 1st lane and the negative control is located on the 9th lane.

The **Figure 4.4** shows two significant and a thin DNA bands on the lane below the wells that are pipetted with the PCR product from A_1 and B_1. This means the application of the PCR profile with 51.4°C of annealing temperature and the PCR mixture as shown in Table 3.1 is relatively efficient for PCR to be carried out on template DNA of JS 1. The PCR product in 50 µL is needed to be prepared in order to get DNA sequence of the extracted DNA from the out-sourced sequencing center. The visualization on the 50 µL PCR product using only DNA template from JS 1 on 1% agarose gel is carried out after the PCR process and it is shown in the **Figure 4.5**.

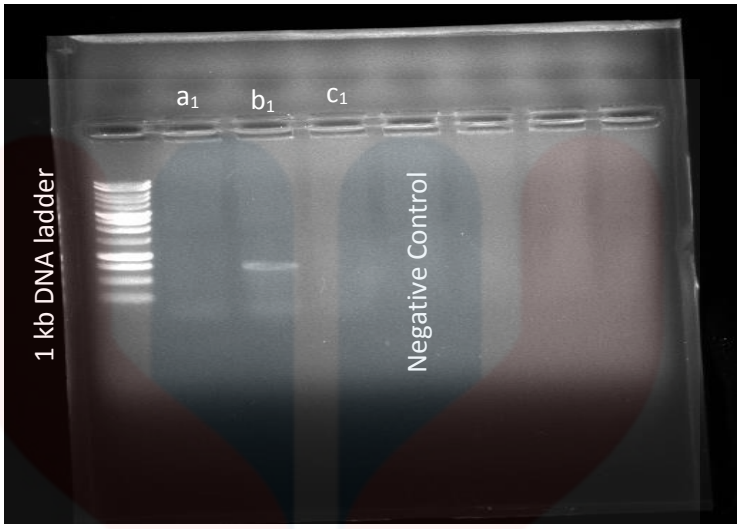


Figure 4.5 shows the result of gel electrophoresis visualization of DNA bands via 1% agarose gel electrophoresis for 50 µL PCR products. A dimming DNA band is shown on the third lane from PCR product sample b_1. 1kb DNA ladder is contained on the 1st lane from left and the negative control is located at the 4th lane.

From **Figure 4.5**, there is only one thin and diming DNA band that can be observed after visualized on 1% agarose gel. There are 3 units of PCR products derived from the PCR reaction mixture using DNA template from JS 1, which is, a_1, b_1, and c_1 respectively in the sequence from left to right. 2 µL of each PCR products were pipetted into the first three wells beside the well pipetted in with 1 kb DNA ladder. Only b_1 shows DNA band in the cell, signifying that B_1 is containing sufficient concentration of DNA to be sent for sequencing by the out-sourced sequencing center.

4.2.2 Primer's dimer

Primer-dimer is a one of the PCR products, which is potentially to be produced every time PCR has been carried out. It is the attachment of primer molecules because of the complementary base strings of primer's 5' ends are still attached to each other after the Taq-polymerase have initiated the annealing and elongation phase (Vallone & Butler, 2004).

Many of the gel electrophoresis resulted in forming dimers that whether the DNA bands are present or not in the gel, it would still be shown inside if the inducing dimer issue in unsolved. The presence of dimer could bring effect to the misinformation in observing DNA bands. It is not a big issue in the case when DNA band can be observed as can be seen in **Figure 4.6**. If the DNA band is not seen and the dimer is observed, then it could be implying issue in the concentration of primer or the quality of the DNA extracted. The presence of primer-dimer might hinder the amplification of the DNA sequence targeted for PCR amplification. There are a few elements that could be affecting the occurrence of primer-dimer other than the primers concentration, including the concentration of magnesium chloride ($MgCl_2$) and dNTP, annealing temperature, and the concentration of template DNA (Bloch, Raymond, & Read, 1994). In this study, the concentration of primers had not change as the primer concentration is set as a constant variable in establishing an efficient PCR profile. There could be various explanation corresponding to the theories mentioned above, including the excessive primer, insufficient dNTP or DNA template, insufficient amount of $MgCl_2$ and etc. However, that is beyond the studies of this study's focus.

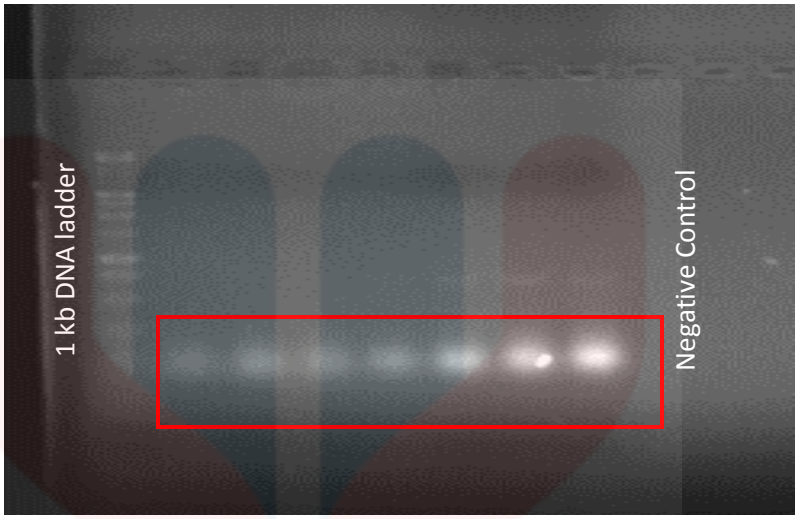


Figure 4.6 shows the sample of primer-dimer situation which extends below the 1kb ladder at the end of each lane after PCR carried out in this study.

4.2.3 Gel Electrophoresis

Gel electrophoresis is an inevitable process to examine the result of DNA extraction and Polymerase Chain Reaction. This process allows the examination of DNA quality to be examined after the extraction and PCR have been carried out. The agarose content proportion was firstly optimized to minimize the possibility of getting DNA bands smeared. The optimization started from 1% agarose gel, where 50g of agarose powder is dissolved in 50ml of 1x TAE buffer added along with 5µl of Redsafe Nucleic acid staining dye. Expectedly, during the visualization of the DNA extracted from the albino squirrel, JS 1, JS 2 and J3, the DNA bands observed are prominent also without dragged and smeared. The agarose concentration has therefore been verified to be usable in visualizing the results of PCR products throughout the research. Agarose gel electrophoresis is used to separate molecules including nucleic acid, based on base-pair counts. The charges of DNA derived from the phosphate group of nucleic acid, makes the DNA travel through

the semi-permeable agarose gel based on the charge given on both end of the electrophoresis tank, where the negative-charged molecules will move towards the positively-charged end (Garner & Revzin, 1981).

However, DNA quality visualization is not convincing without the running along with the DNA ladder. DNA ladder is basically a DNA sample that have known number of base pairs in it which can act as standardized positive control. This allows the DNA fragment's base pair number to be easily identified by comparing on the standard base pair count of the DNA ladder. The DNA ladder used in carrying out gel electrophoresis is 1 kb, which means it contain fragments that have base pair counts from 1 to 1000. The advantage of applying DNA ladder shows up more when targeted DNA fragments of PCR product has carried out gel electrophoresis. It is important to know whether the band showed up at the right base-pair standard comparing to DNA ladder in order to find out whether the DNA band showed up is the targeted fragment or not. The DNA ladder is used to differentiate whether the PCR product is the targeted DNA segment or not. This can be done by referring the position of the DNA bands observed along the lane with the DNA ladder's bands that can represent different number of base pairs that serves as a scale during separation via electrophoresis. In this case, the base pair number of the targeted segment of DNA from COX1 gene is observed to be located around the position at approximately 625-650bp.

4.3 Phylogenetic Analysis (Neighbor Joining)

The result of this study is concluded by the generated phylogenetic tree. As shown in the result **Figure 4.7**, the phylogenetic analysis method that has been chosen to infer the evolutionary history of the albino squirrel with other comparative species is the Neighbor-Joining method (Saitou and Nei, 1987). The neighbor joining method works as a bottom-up analyzing method for nucleotide data that cluster up the highly similar nucleotide data species together in a taxa and present it in the form of phylogenetic tree. The similarities of each individual's nucleotide data are represented by a numerical figure calculated through algorithms provided by all kinds of bioinformatics soft wares and other similar function calculator. The nucleotide data of each individual is then calculated repetitively along with other individual nucleotide data to get the overall "dissimilarities" between each individual that is calculated using the input of distance matrix. By calculating the dissimilarities in the distance matrix, the distance of each unit of individuals will be able to resolve and are able to be plotted on the phylogenetic tree. Hence, a phylogenetic tree of Neighbour-Joining model for the identification of albino squirrel is synthesized in **Figure 4.7** and **Figure 4.8**.

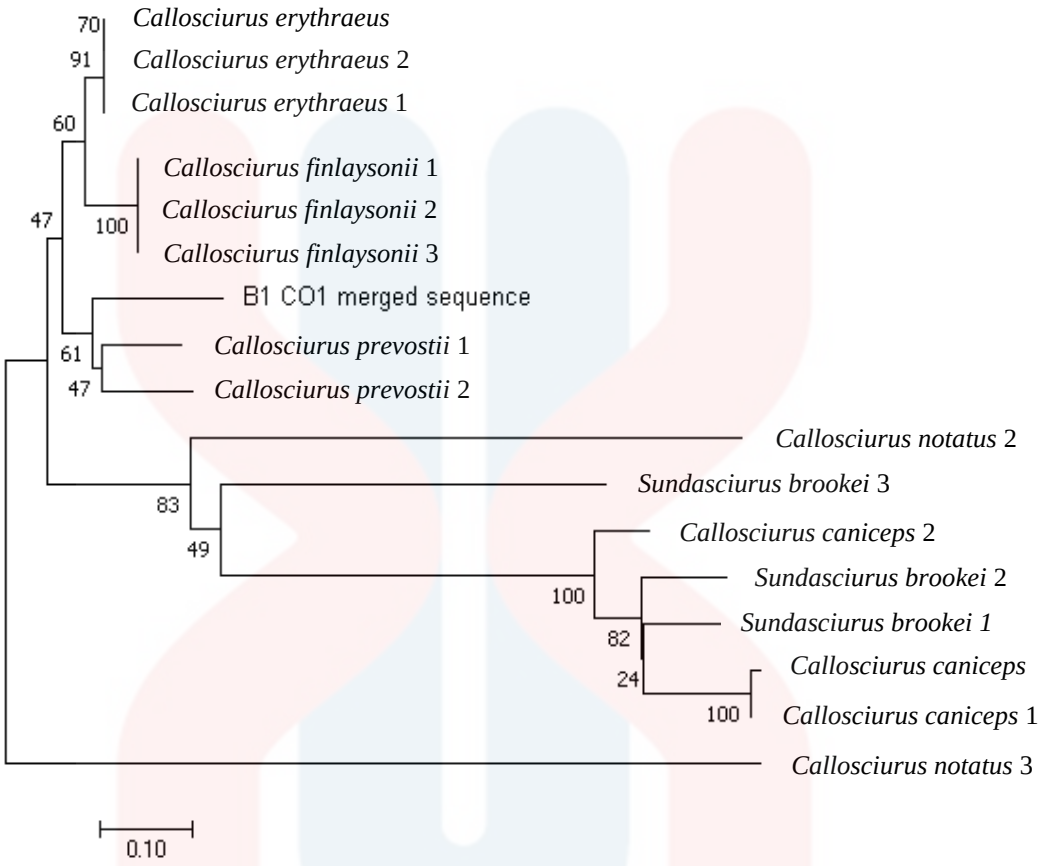


Figure 4.7 shows the molecular phylogenetic analysis on sequence of BLAST matches by neighbor-joining tree computed using the Kimura-2 parameter model bootstrapping in 10,000 replicates

From **Figure 4.7**, the phylogenetic tree shows that the merged sequence of B_1 is analyzed through the Neighbor joining method along with other sequences from the species that is listed top in the hits of the B_1 BLAST result as shown in the **Figure 4.8**. There are 3 individual's nucleotide sequence of the COX1 gene included in each species to form the Neighbour joining tree. The result of the analyzation resulted in where the sequences of B_1 is located in between the cluster groups *of Callosciurus prevostii* and *Callosciurus finlaysonii*. However, the subject sequence appears to share the same clade with *Callosciurus prevostii* and located nearer to it in the phylogenetic tree.

Table 4.2 shows the list of matches of the B_1 merged sequence BLAST result on NCBI.

Sequences producing significant alignments	Max Score	Total Score	Query cover (%)
Callosciurus erythraeus mitochondrion, complete genome	581	1103	85
Callosciurus notatus strain ATCC CRL-1926 cytochrome oxidase subunit 1 (COI) gene, partial cds; mitochondrial	547	547	88
Callosciurus notatus isolate 3 mitochondrion, complete genome	516	516	92
Callosciurus notatus isolate 30 mitochondrion, complete genome	532	532	92
Callosciurus sp. 2 AG-2015 mitochondrial partial CO1 gene for cytochrome oxidase subunit 1, specimen voucher MIBZPL 07392, halophyte CXH11	532	1047	90
Prosciurillus weberi voucher MZB:6255 mitochondrion, partial genome	483	483	91
Prosciurillus weberi voucher MZB:6354 mitochondrion, partial genome	479	479	91
Sundasciurus brookei isolate 6 mitochondrion, complete genome	515	1001	88
Callosciurus prevostii isolate 8 mitochondrion, complete genome	506	987	87
Callosciurus prevostii voucher ROM:102221 cytochrome oxidase subunit 1 (COI) gene, partial cds; mitochondrial	506	987	87

From the filtered and related-species that is shown in **Figure 4.7**, a few species individuals of squirrel from different genus distributed in Indo-Malayan region are once again put together and to be analyzed to synthesize the phylogenetic tree as shown in **Figure 4.8**.

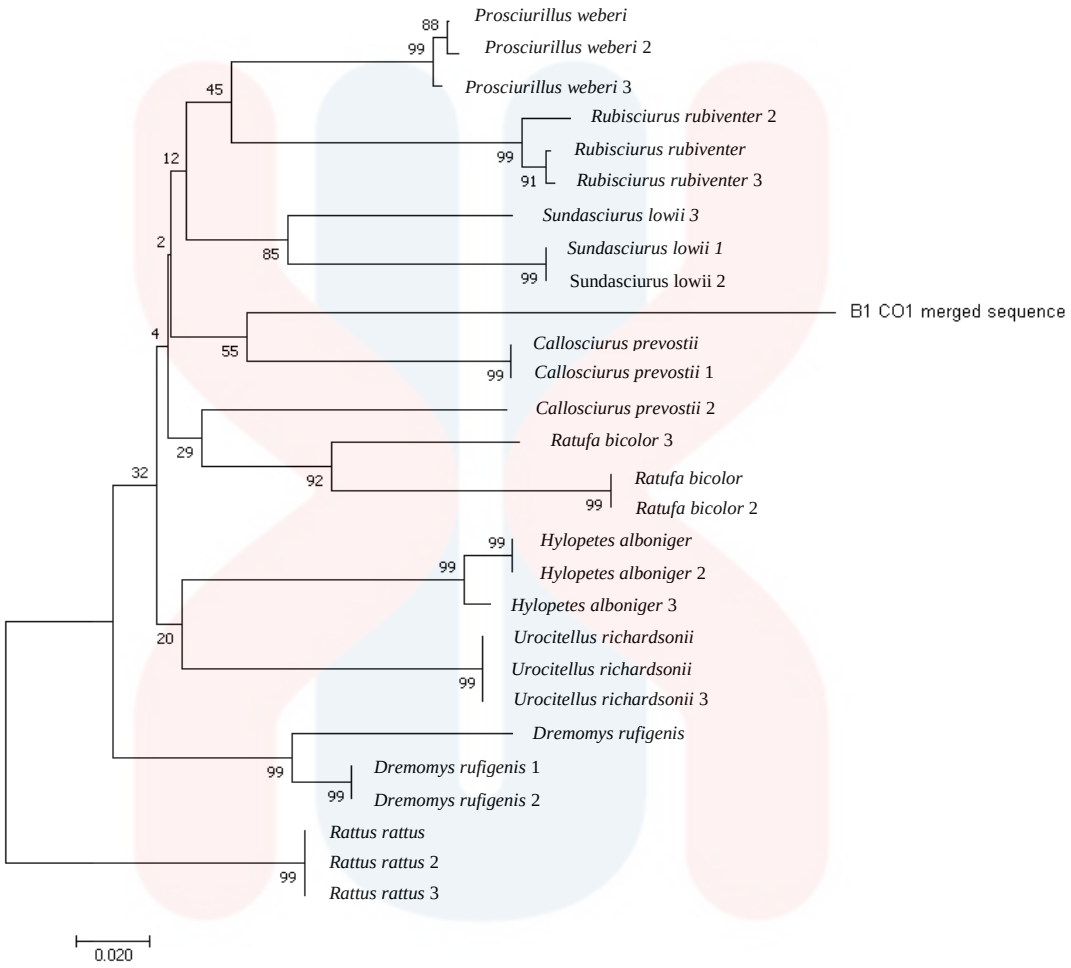


Figure 4.8 shows the molecular phylogenetic analysis by neighbor-joining tree across genus of Sciuridae in Indo-Malayan region computed using the Kimura-2 parameter model bootstrapping in 10,000 replicates.

Figure 4.8 is a phylogenetic tree that is generated using DNA barcode of the PCR product of the albino squirrel tissue sample, B_1. From the analysis result along with 27 other nucleotide sequences from Indo-Malayan region Sciuridae family's genus sequence samples labelled as above, the sequence of B_1 merged sequence located at the end of an isolated branch between the group that is made up sequences of *Callosciurus prevostii* and *Sundasciurus lowii*. There are 3 individual's nucleotide sequence of the COx1 gene included in each species to form the Neighbour Joining tree. The **Figure 4.8** also displays

the frequencies of similarities in terms of distance matrix beside each branch generated from the phylogenetic tree above. The bottom right corner of the figure is attached with the scale of the branch length in this figure. The result that can be observed from this figure is that the B_1 sequence is not clustered in any of the taxa group available and it is prominently branched away longer than the other sequences

From the result of bioinformatic analysis software, MEGA 7, the total branch length which infers the changes in sites or positions is up to 304. Respective to that, the sum of branch length is up to 1.28165293. The revolved locus of the analysis clusters is towards the nucleotide sequences of the novel mitochondrial gene, COX1. In regards of these variables of the phylogenetic tree, the analysis shows that the merged sequence of B_1 is located in isolation between the cluster of the *Sundasciurus lowii* group and the *Callosciurus prevostii* group. Each of the groups or clusters are formed by three individual's nucleotide sequences respectively to the species named on each branch as shown in **Figure 4.7** and **4.8**. By comparing the phylogenetic trees of **Figure 4.7** and **4.8**, the analysis escalated from within the matches of B_1 BLAST results towards the scale of Indo-Malayan region Sciuridae genus individuals which functions similar to the concept of bottom-up in order to filter out the possibilities that species other than those paired with in **Figure 4.8** can fit in.

The sequence of B_1 is inferred as resolved among the comparative sequences due to the result of being in the same clades that is closely placed together clustered in the species group of *Callosciurus prevostii* in both the phylogenetic tree in **Figure 4.7** and **4.8**. Furthermore, the comparison in **Figure 4.7** and **4.8** shows that the B_1 sequence still remain comparatively close with the cluster group of *Callosciurus prevostii*. This species

sequence is also one of the matches in the B_1's BLAST result that possess the identical level up to 81% (refer to **Table 4.2**）.

The result of the analysis proves the hypotheses that B_1 is identical to *Calloscirus prevostii* or can be said highly possible to be *Callosciurus prevostii*, but it is certain that the species of B_1 is relatively closer in terms of nucleotide sequence similarity compare to other species listed in the phylogenetic tree in **Figure 4.7** and **4.8**. It is also noticeable that the species *Rattus rattus* is located at the root of branches in the phylogenetic tree amongst different genus individuals in **Figure 4.8**. This is because *Rattus rattus*, also known as house rat or black rat is most common and widely spread rodents across continents due to their high adaptative ability (Pye, Swain, & Seppelt, 1999). By locating the sequence of *Rattus rattus* at the root of the tree branches of phylogenetic analysis, it infers the presumption that all the individuals across the multiple genus in Sciuridae family is all from the order of Rodentia. These information in total is currently suffice to resolve the species identity of the albino squirrel. The findings based on the objectives of this study is therefore determined as a positive result.

There are hardships in obtaining a good DNA sequence alignment quality. It directly affects the pair wise distance which in turn alternating the calculated values in the distance matrix. When the distance value of the sequenced nucleotides of B_1 in distance matrix cannot represent the actual value of the albino squirrel. Of course, the technologies in recent decades has developed to the level where there are algorithms created that enables the masking of defected sequence that has caused the inefficiency in pairwise distance generation.

However, this convenience is accompanied with the of risk of removing considerable amounts of nucleotides data along when the defected pairwise distance data is filtered. In consideration of the unstable quality nucleotide sequence of B_1, it is preferably to not filter out the possibly useful nucleotide data that can form genetic pairwise distance. Therefore, the sequence that is used to carry out the phylogenetic analysis remain as it is synthesized and caused for the low bootstrap number.

CHAPTER 5

CONCLUSION & RECOMMENDATION

In conclusion, the sequence obtained from the albino squirrel's mitochondrial DNA defined to be *Callosciurus prevostii* based on the inferration in the phylogenetic tree. This is due to the position of the albino squirrel's DNA sequence that is analyzed in both Neighbour-Joining tree is within the clades of COX1 DNA sequence of *Callosciurus prevostii*. The results are also showing *Callosciurus prevostii* is the species that has the closest distance in terms of nucleotide sequences data. Hence, the actual species of the albino squirrel has resolved from the result of the phylogenetic analysis with a high possibility to be *Callosciurus prevostii.*

Regarding the result of the study, there are some recommendations that may improve the results obtained. Firstly, improving the study and research towards the subject of Sciuridae. This can help to ease out the identification of the albino squirrel by filtering out the species listed in BLAST nucleotide sequence match, by characterizing the albino squirrel morphological study and geographic distribution. From that, the species that will be shortlisted will possess more characteristic matches before carrying out the phylogenetic analysis on its mitochondrial gene.

Other than that, it is also recommendable to improve the understanding in carrying out PCR and its related studies. The unproficiency in carrying out optimization and the actual PCR has a lot of impairments in the result obtained. This has also caused a lot of delayed progress for thesis writing and PCR product sequencing due to repetitive attempt to get desired results of PCR through 1% agarose gel electrophoresis. Therefore, it is very recommendable to apply more knowledge and studies in carrying out PCR.

REFERENCE

Arnaud, F., & Bamber, R. . (1989). The biology of Pycnogonida, Advance Marine Biology 24:1-95

Austerlitz, F., David, O., Schaeffer, B., Bleakley, K., Olteanu, M., Leblois, R., Laredo, C. (2009). DNA barcode analysis: a comparison of phylogenetic and statistical classification methods. *BMC Bioinformatics*, *10*(Suppl 14), S10. https://doi.org/10.1186/1471-2105-10-S14-S10

Bloch, W., Raymond, J. C., & Read, A. R. (1994). Compostions and methods for inhibiting dimerization of primers during storage of Polymerase Chain Reaction reagents. *United States Patent*, (19).

Bohmann, K., Evans, A., Gilbert, M. T. P., Carvalho, G. R., Creer, S., Knapp, M., de Bruyn, M. (2014). Environmental DNA for wildlife biology and biodiversity monitoring. *Trends in Ecology and Evolution*, *29*(6), 358–367. https://doi.org/10.1016/j.tree.2014.04.003

Carden, S. M., Boissy, R. E., Schoettker, P. J., & Good, W. V. (1998). Albinism: Modern molecular diagnosis. *British Journal of Ophthalmology*, *82*(2), 189–195. https://doi.org/10.1136/bjo.82.2.189

Cenis, J. L. (1992). Rapid extraction of fungal DNA for PCR amplification. *Nucleic acids research*, *20*(9), 2380.

Davis W. Ronald, Thomas, M., Cameron, J., Scherer, S., Thomas, P. S. J., Stewart, S., & Padgett Richard A. (1980). The ability to rapidly isolate DNA from a large number of individual organisms has greatly facilitated the characterization of the genome of. In *Methods in Enzymology* (Vol. 65, pp. 404–411).

Delsuc, F., Brinkmann, H., & Philippe, H. (2005). Phylogenomics and the reconstruction of the tree of life. *Nature Reviews Genetics*, *6*(5), 361–375. https://doi.org/10.1038/nrg1603

Dennerlein, S., & Rehling, P. (2015). Human mitochondrial COX1 assembly into cytochrome c oxidase at a glance. *Journal of Cell Science*, *128*(5), 833–837. https://doi.org/10.1242/jcs.161729

Edition, E., Campbell, N., & Reece, J. (2008). Chapter 26 Phylogeny and the Tree of Life, *6*, 1–54.

Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., ... & Bibillo, A. (2009). Real-time DNA sequencing from single polymerase molecules. *Science*, *323*(5910), 133-138.

Enders, C. K. (2005). Estimation by maximum likelihood. *Encyclopedia of Behavioral Statistics*, (Ml), 1164–1170.

Faloona, F., & Mullis, K. (1987). Specific synthesis of DNA in vitro via a polymerase catalysed chain reaction. In *Methods in Enzymology* (pp. 335–350).

Farris, J. S. (1970). Methods for computing Wagner tress. *Systematic Zoology*, *19*(1), 83–92.

Felsenstein, J. (1985). Confidence limits on phylogenetic: an approach using the bootstrap. *Evolution*, *39*(4), 783–791.

Freckleton, R. P., Harvey, P. H., & Pagel, M. (2002). Phylogenetic analysis and comparative data: a test and review of evidence. *The American Naturalist*, *160*(6), 712-726.

Garner, M. M., & Revzin, A. (1981). A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the *Escherichia coli* lactose operon regulatory system. In *Nucleic Acids Research* (Vol. 9, pp. 3047–3060).

Gaut, B. S., & Lewis, P. O. (1995). Success of maximum likelihood phylogeny inference in the four-taxon case. *Molecular Biology and Evolution*, *12*(1), 152-162.Glasel, J. A. (1995). Validity of nucleic acid purities monitored by 260nm/280nm absorbance ratios. *Biotechniques*, *18*(1), 62-63.

Haines, A. M., Tobe, S. S., Kobus, H. J., & Linacre, A. (2015). Properties of nucleic acid staining dyes used in gel electrophoresis. *Electrophoresis*, *36*(6), 941-944.

Hebert, P. D. N., Ratnasingham, S., & de Waard, J. R. (2003). Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society B: Biological Sciences*, *270*(Suppl_1), S96–S99. https://doi.org/10.1098/rsbl.2003.0025

Huang, X., & Brule-Babel, A. (2012). Sequence diversity , haplotype analysis , association mapping and functional marker development in the waxy and starch synthase IIa genes for grain-yield-related traits in hexaploid wheat (Triticum aestivum L .), 627–645. https://doi.org/10.1007/s11032-011-9649-8

James, V., Maria, S., Frans.J, D., & Lupski, J. R. (1994). Genomic Fingerprinting of Bacteria Using Repetitive Sequence-Based Polymerase Chain Reaction. In *Methods in Molecular And Cellular Biology* (pp. 25–40).

Kimura, M. (1980). Journal of Molecular Evolution ©, *16*(1330), 111–112.

Kirby, K. S. (1956). A New Method for the Isolation of Ribonucleic Acids from Mammalian Tissues Insoluble protein. *Biochemical Journal*, *64*(1953), 405–408.

Kress, W. J., Wurdack, K. J., Zimmer, E. A., Weigt, L. A., & Janzen, D. H. (2005). Use of DNA barcodes to identify flowering plants. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(23), 8369–8374. https://doi.org/10.1073/pnas.0503123102

Kumar, S., Stecher, G., Tamura, K., & Medicine, E. (2016). MEGA7 : Molecular Evolutionary Genetics Analysis version 7 . 0 for bigger datasets, 1–11.

Kumaran, V. J (2009). The Phylogenetic Relationships of Megachiroptera In Malaysia Inferred From Morphological and DNA Analyses. *Universiti Malaysia Sarawak, Faculty of Resource Science and Technology, Sarawak, Malaysia.*

Laake, P., Benestad, H. B., & Olsen, B. R. (Eds.). (2007). *Research methodology in the medical and biological sciences*. Academic Press.

Lee, S.T., Nicholis, R. D., Phil, D., Bundey, S., Laxova, R., Musarella, M., & Spritz, R. A. (1994). The New England Journal of Medicine

Luo, A., Zhang, A., Ho, S. Y. W., Xu, W., Zhang, Y., Shi, W., Zhu, C. (2011). Potential efficacy of mitochondrial genes for animal DNA barcoding: A case study using eutherian mammals. *BMC Genomics*, *12*. https://doi.org/10.1186/1471-2164-12-84

Mattos, C., & Ringe, D. (2001). Proteins in organic solvents, 9–11.

Meyer, C. P., & Paulay, G. (2005). DNA barcoding: Error rates based on comprehensive sampling. *PLoS Biology*, *3*(12), 1–10. https://doi.org/10.1371/journal.pbio.0030422

Nedyalkov, N., Koshev, Y., Raykov, I., & Bardarov, G. (2014). Color variation of small mammals's (Mammalia: Rodentia and Insectivora) coats from Bulgaria. *North-Western Journal of Zoology*, *10*(2), 314–317.

Nozaki, Y., & Tanford, C. (1971). The Solubility of Amino in Aqueous Ethanol Acids and Two Glycine Dioxane Solutions Peptides. *The Journal of Biological Chemistry*, *246*(April 10), 2211–2217.

Park, D. S., Suh, S. J., Oh, H. W., & Hebert, P. D. N. (2010). Recovery of the mitochondrial COI barcode region in diverse Hexapoda through tRNA-based primers. *BMC Genomics*, *11*(1). https://doi.org/10.1186/1471-2164-11-423

Pye, T., Swain, R., & Seppelt, R. D. (1999). Distribution and habitat use of the feral black rat (Rattus rattus) on subantarctic Macquarie Island. *Journal of Zoology*.

Roland.R Rueckert, & Dunker, A. K. (1969). Observations on Molecular on Polyacrylamide Gel Weight, (18), 5074–5081.

Roux, K. H. (2009). Optimization and Troubleshooting in PCR Optimization and Troubleshooting in PCR. *PCR Primer: A Laboratory Manual*. https://doi.org/10.1101/pdb.ip66

Sabnis, R. W. (2010). *Handbook of biological dyes and stains: synthesis and industrial applications*. John Wiley & Sons.

Saitou, N., & Nei, M. (1987). The neighbour-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evo*, *4*(4), 406–425. https://doi.org/citeulike-article-id:93683

Savolainen, V., Cowan, R. S., Vogler, A. P., Roderick, G. K., & Lane, R. (2005). Towards writing the encyclopaedia of life: an introduction to DNA barcoding. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1462), 1805–1811. https://doi.org/10.1098/rstb.2005.1730

Seifert, K. A., Samson, R. A., deWaard, J. R., Houbraken, J., Levesque, C. A., Moncalvo, J.-M., … Hebert, P. D. N. (2007). Prospects for fungus identification using CO1 DNA barcodes, with Penicillium as a test case. *Proceedings of the National Academy of Sciences*, *104*(10), 3901–3906. https://doi.org/10.1073/pnas.0611691104

Sivakumar, K., Anandan, R., Muthupriya, P., Gopikrishna, M., & Altaff, K. (2013). Phylogenetic analysis of thermocyclops decipiens with reference to 18S rDNA. *Indian Journal of Science and Technology*, *6*(12), 5585–5592.

Smith, M. A., Poyarkov, N. A., & Hebert, P. D. N. (2008). CO1 DNA barcoding amphibians: Take the chance, meet the challenge. *Molecular Ecology Resources*, *8*(2), 235–246. https://doi.org/10.1111/j.1471-8286.2007.01964.x

Sober, E. (1983). PARSIMONY IN SYSTEMATICS: Philosophical Issues. *Ann. Rev. Ecol. Syst*, *14*, 335–357. https://doi.org/10.1146/annurev.es.14.110183.001525

Thompson, D., Seligmann, B. E., & Gordon, D. A. (2014). *U.S. Patent Application No. 14/266,884*.

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., & Higgins, D. G. (1997). The CLUSTAL X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research*, *25*(24), 4876–4882. https://doi.org/10.1093/nar/25.24.4876

Tripathy, S. K., Maharana, M., Ithape, D. M., Lenka, D., Mishra, D., Prusti, A., & Raj, K. R. (2017). Exploring Rapid and Efficient Protocol for Isolation of Fungal DNA. *Int. J. Curr. Microbiol. App. Sci*, *6*(3), 951-960.

Valentini, A., Pompanon, F., & Taberlet, P. (2009). DNA barcoding for ecologists. *Trends in Ecology and Evolution*, *24*(2), 110–117. https://doi.org/10.1016/j.tree.2008.09.011

Vallone, P. M., & Butler, J. M. (2004). AutoDimer : a screening tool for primer-dimer and hairpin structures, *37*(2).

Van Houdt, J. K. J., Breman, F. C., Virgilio, M., & De Meyer, M. (2010). Recovering full DNA barcodes from natural history collections of Tephritid fruitflies (Tephritidae, Diptera) using mini barcodes. *Molecular ecology resources*, *10*(3), 459-465.

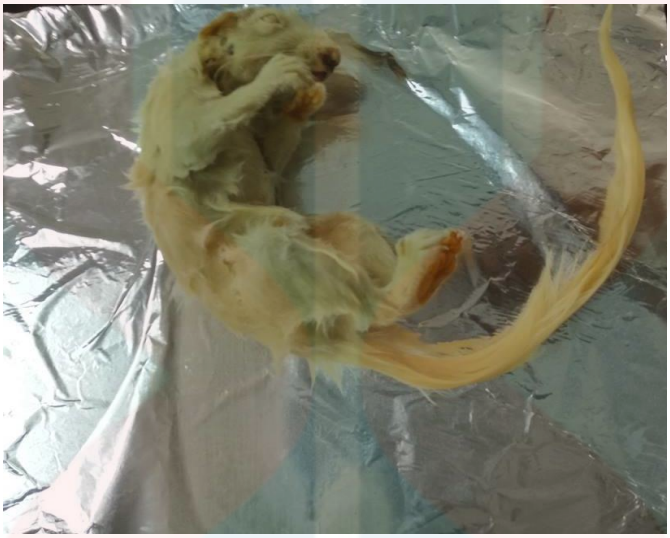
Veena, S., Thomas, S., Raje, S. G., & Durgekar, R. (2011). Case of leucism in the spadenose shark, Scoliodon laticaudus (Müller and Henle, 1838) from Mangalore, Karnataka. *Indian Journal of Fisheries*, *58*(1), 109–112.

Vrijenhoek, R. (1994). DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol Mar Biol Biotechnol*, *3*(5), 294-9.

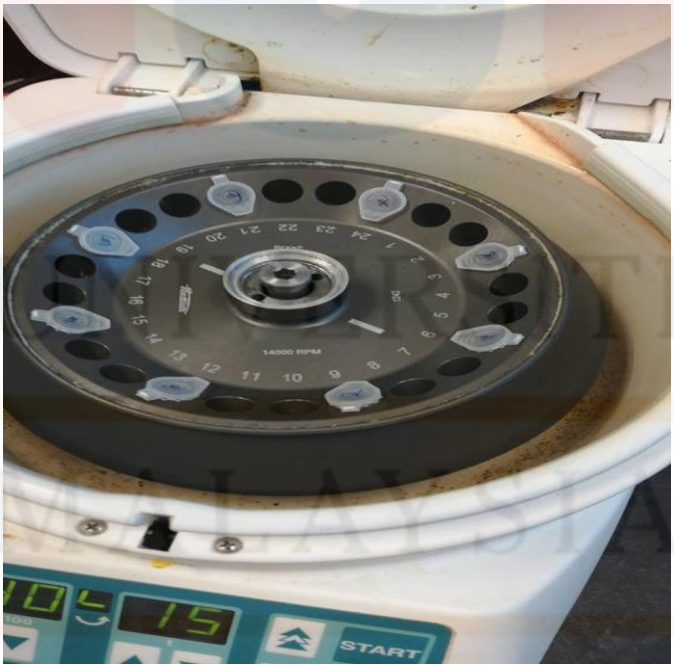
Wassarman, Paul, M., & Kornberg, R. D. (1990). Methods in Enzymology. In *Molecular cloning: a laboratory manual*. *Analytical Biochemistry* (Vol. 170, pp. 182–183). San Diego, CA. https://doi.org/10.1016/0003-2697(90)90595-z

Woese, C. R. (2000). Interpreting the universal phylogenetic tree. *Proceedings of the National Academy of Sciences*, *97*(15), 8392–8396. https://doi.org/10.1073/pnas.97.15.8392

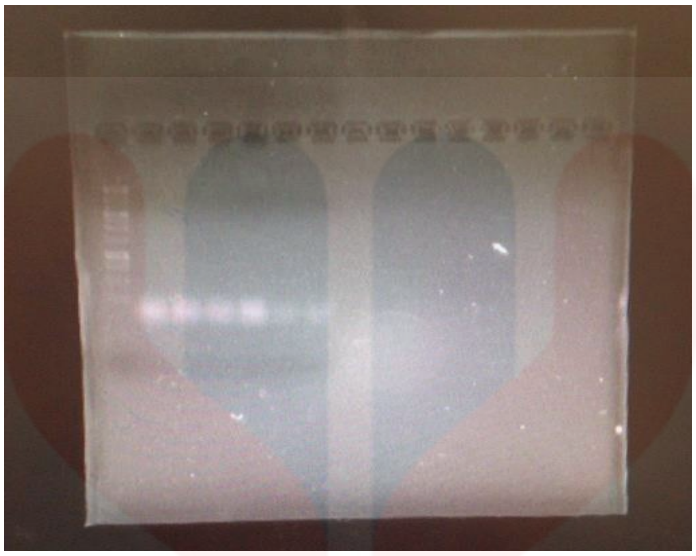
APPENDICES



The picture of the albino squirrel before dissected for its liver tissue.

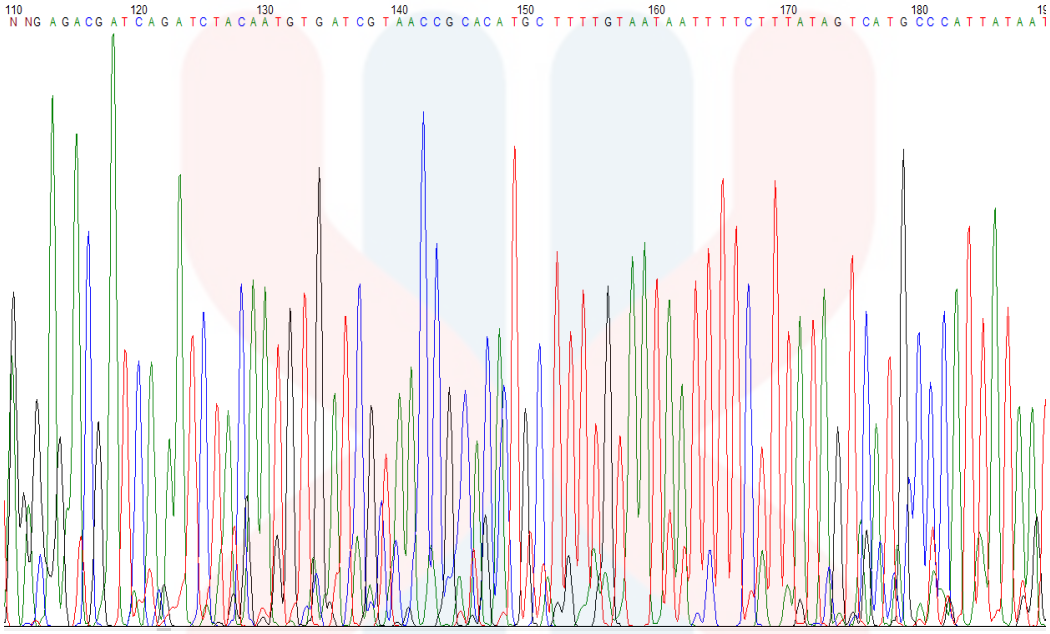


The picture of micro-centrifugation machine with acknowledged microcentrifugation tube location.



Primer dimers observed in 1% agarose gel electrophoresis visualization on one of the failed PCR products.

PCR product that shows inconspicuous DNA bands of the PCR product when visualized on 1% agarose gel electrophoresis



The chromatogram profile of the sequence B_1 generated from the PCR product sent for sequencing.

FYP 1	23rd March – 15th April 2018	Completion of Chapter 1
		Completion of Chapter 2
		Completion of Chapter 3
	10th April 2018	Research Proposal Submission
	Lab work	
FYP 2	15th July 2018	DNA Extraction
	17th August 2018	DNA Extraction optimization
	25th September 2018	PCR annealing temperature
	15th October 2018	PCR amplification of targeted segment
	14th November 2018	DNA Sequencing from 50 µL PCR product
	29th November 2018	Phylogenetic Analysis
	Report Writing	
	20th August 2018	Chapter 1, 2 and 3 revised
	30th November 2018	Chapter 4 completed
	4th December 2018	Chapter 5 completed
	10th December 2018	Thesis submission

Chemical used

CTAB solution

Proteinase K

Chloroform

Isoamyl alcohol

70% ethanol

NaCl solution

Deionized distilled water, ddH_2O

TE buffer

Go Flexi-Taq PCR buffer

TAE buffer

1kb DNA ladder

Sodium acetate

Absolute ethanol molecular grade

dNTP

Taq DNA polymerase

$MgCl_2$

Universal COx1 Barcoding Primers LCO1490, HCO2198

Agarose powder

Redsafe nucleic acid staining solution