GENERATING SMALL MAMMALS MITOGENOME REFERENCE DATASET FOR MALAYSIAN SPECIES

PUTERI NUR SYAHZANANI BINTI JAHARI

A thesis submitted in fulfilment of the
requirements for the award of the degree of
Master of Philosophy

Faculty of Science
Universiti Teknologi Malaysia

OCTOBER 2019

DEDICATION

This thesis is dedicated to beloved family and supervisor who taught me that the best kind of knowledge to have is that which is learned for its own. Also, throughout the journey, I've been taught that even the largest task can be accomplished if it is done one step at a time.

ACKNOWLEDGEMENT

% L V P L O O D K «

My special thanks go to my supervisor, Dr Faezah Mohd Salleh. She has taught me many things, not only about this project but also about how to live amazing life in this world. Without her persistence and perseverance in guiding me throughout this journey, I could not have finished. Also, this project might not be completed without her faith in me to do it in Copenhagen. Thank you for every amazing opportunity you have given to me. Since I started this project, too many things have happened in my life, and it was not easy. Thanks to you for always being my greatest support system be it in knowledge transfer, survival skills, emotional support and financial support. Also, I would like to express my deepest gratitude to my co-supervisor, Prof Mohd Shahir Shamsir. His support, guidance and encouragement have been with me since the day I left to Kent to do FYP project in 2016, which is way before this project was started and until now, I still have his endless support and motivation.

They are so many people in Copenhagen that I would like to thank for making this project possible. First, I am indebted to Tom Gilbert for always making time for the Skype meeting to discuss this project despite his busy schedule. Also, I am grateful to have him that always welcomed me with an open arm to experience the working culture in his established lab. My thanks go to Stine Richter who help me in the lab and Ashot, who always answered to me when I in the middle of complication during running the mitogenome assembly. Thanks to my lovely friend, Catia who always support and provide her shoulder for me to cry on. My life would not be so cheerful in Copenhagen without you. Next, of course, to the families (Kak 3 H N M D K ¶ V D Q G . D N (U L Q ¶ was the key, thank you for making my Z K L O H , life so wonderful.

My appreciation also extends to the staff, seniors and friends in PBT and UCPH lab for assistance, kind advice and supports throughout this journey.

My amazing life would not be so amazing without this wonderful friend, Hanem. Thank you for always creating joy during this journey and would always give support whenever I am in sorrow. I am so grateful to have you standing by me.

I am dedicating this thesis to my family members, Ibu, Abah, Abgngah, Abg Ishi, Adik wani who would have been so proud of everything I do. Thank you for their endless support and for always believing in me. To my soulmate, Nik Ahmad Faiz, thank you for the motivation and enthusiasm that has kept me going for these past years.

I would not be where I am today without all of you. This journey has left a big impact (in a good way) in my life that I would never forget. Thank you everyone!

ABSTRACT

0DODV\VLD¶V EL RORJLFDO GLYHUVLW\ rapidly DPRQJ WK declining due to various human activities and climate change. Despite the continuous loss in biodiversity with the most recent death of our last Sumatran Rhino in May 2019, Malaysia's biodiversity data is still poorly characterized and systematically documented. Most species data have limited visibility with very abysmal publications which are mostly restricted to morphological traits and lack genomic data. As in line with the coQ W U \¶V 1D W L RQ D O 3R O L F \ R-Q %L R O R J 2025 to combat biodiversity loss and global effort to sequence all life by 2028 (Earth Biogenome Project), this study has focused on generating Malaysian small mammals mitogenome reference dataset. The genomic DNA of two fresh tissue samples (Balionycteris maculata and Callosciurus notatus) were extracted, fragmented to 300 bp, and further constructed into Illumina-compatible libraries using BEST protocol. Next, about 15 amplification cycles during library indexing was used to produce maximum data output with high complexity, and low clonality suggesting the rare variants could be easily detected. Prior to sequencing using BGISEQ-500 platform, the 3-indexed libraries (including extraction blanks) with approximately 300-400 bp were pooled to equimolar DNA (<12,000 pmol/L). Approximately 5 gigabases raw sequence data per sample were generated comprising of the whole genome data (mitochondrial DNA and nuclear DNA). The new high quality mitogenomes has been successfully assembled using MITOBIM and PALEOMIX with an average size of 16-17 Kbp and an average depth of coverage of 274. The detailed pipeline and challenges on mitogenome assembly for species with and without reference genome in Genbank was discussed. The mitogenomes were further annotated to its 37 designated genes via MitoZ. The robust pipeline of mitogenome sequence generation established in this work could be applied to generate more genomic data from thousands of tissue samples available from local biodiversity key players such as Perbadanan Taman Negara Johor, FRIM and PERHILITAN. Further enrichment of DNA reference database will strongly magnify species detection in invertebrate-derived DNA (iDNA) research for biodiversity assessment, wildlife forensics to monitor illegal trade of endangered species in this region, as well as population genetic studies.

ABSTRAK

Kepelbagaian biologi Malaysia adalah antara yang terkaya di dunia tetapi kian merosot disebabkan pelbagai aktiviti manusia dan perubahan iklim yang kerap berlaku. Walaupun menghadapi kehilangan biodiversiti yang berterusan dengan kematian terbaru Badak Sumbu Sumatra terakhir pada bulan Mei 2019, data biodiversiti di Malaysia masih kurang didokumentasikan secara sistematik. Kebanyakan data spesies terhad hanya kepada penerbitan yang tidak teratur. Malah, kebanyakannya terhad kepada sifat morfologi dan kekurangan data genom. Selaras dengan Dasar Kebangsaan mengenai Kepelbagaian Biologi Negara 2016-2025 untuk memerangi kehilangan biodiversiti dan usaha global untuk menjejaki semua hidupan menjelang 2028 (Projek Biogenom Bumi), kajian ini memberi tumpuan kepada penjanaan data rujukan mitogenom mamalia kecil di Malaysia. DNA genomik dua sampel tisu (*Balionycteris maculata* dan *Callosciurus notatus*) diekstrak, dipecahkan kepada 300 bp, dan dibina semula menjadi perpustakaan yang komprehensif. Seterusnya, kira-kira 15 kitaran penguat semasa pengindeksan perpustakaan digunakan untuk menghasilkan jumlah data yang maksimum dengan kerumitan tinggi, dan klonalan yang rendah bagi memudahkan pencarian varian yang jarang dapat dikesan dengan mudah. Kemudian, perpustakaan yang telah diindekskan memanjang kepada kira-kira 300-400 bp, seterusnya dikumpulkan kepada DNA equimolar (<12,000 pmol/L) sebelum dijujukan menggunakan platform BGISEQ-500. Kira-kira 5 gigabase data mentah telah dijana bagi setiap sampel yang terdiri daripada keseluruhan genom (DNA mitokondria dan DNA nuklear). Mitogenom baru yang berkualiti tinggi telah berjaya dihasilkan menggunakan MITOBIM dan PALEOMIX dengan saiz purata 16-17 Kbp dan kedalaman liputan purata 140.27x. Pautan yang terperinci mengenai perhimpunan mitogenom untuk spesies menggunakan genom rujukan dan tanpa genom rujukan di Genbank dibincangkan. Mitogenom ini selanjutnya dikelaskan kepada 37 gen yang ditetapkan melalui MitoZ. Saluran pergerakan urutan mitogenom yang teguh yang ditubuhkan dalam karya ini boleh digunakan untuk menghasilkan lebih banyak data genomik daripada beribu-ribu sampel tisu yang terdapat dalam simpanan badan biodiversiti tempatan seperti Perbadanan Taman Negara Johor, FRIM dan PERHILITAN. Pengayaan pangkalan data rujukan DNA akan membesarkan pengesanan spesies dalam penyelidikan DNA (iDNA) berkenaan dengan penilaian biodiversiti, forensik hidupan liar untuk memantau perdagangan haram spesies terancam di rantau ini, serta kajian genetik populasi.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AULC	-	Amount used for library construction
BEST	-	Blunt-end Single-tube
BGI	-	Beijing Genome Institute
BLAST	-	Basic Local Alignment Search Tool
CBD	-	Convention of Biological Diversity
COI	-	Cytochrome *c* oxidase subunit I
Cytb	-	Cytochrome *b*
DNA	-	Deoxyribonucleic acid
D-loop	-	Displacement loop
EBP	-	Earth Biogenome Project
GOLD	-	Genome Online Database
HPC	-	High Performance Computing
HTS	-	High Throughput Sequencing
iDNA	-	Invertebrate-derived DNA
IUCN	-	International Union for Conservation of Nature
NCBI	-	National Centre for Biotechnology Information
NGS	-	Next Generation Sequencing
qPCR	-	Quantitative polymerase chain reaction
SPRI	-	Solid-phase reversible immobilization

LIST OF SYMBOLS

bp	-	basepair
°C	-	Degree Celcius
Kb	-	Kilobase
µL	-	Microlitre
µM	-	Micromolar
M	-	Molar
ng	-	nanogram
%	-	Percent
rpm	-	Rotary per minute
pg/µL	-	Picogram/microlitre
ng/µL	-	Nanogram/microlitre
PE	-	Paired-end
SR	-	Single-read
Gb	-	Gigabyte
pmol/L	-	picomol/litre

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Background of study

Malaysia, a megadiverse country boasts a myriad faunal community, comprising of 307 mammals; 785 birds; 242 amphibians; 567 reptiles; 470 freshwaters; and 1,400 marine fishes [1]. In spite of having an immense variety of fauna, Malaysia has lost battles on conservation efforts of some native and iconic species due to the country's development pressures and excessive exploitation by poachers [2,3]. Some critically affected species include the Malayan tiger (*Panthera tigris jacksoni*), Sumatran rhinoceros (*Dicerorhinus sumatrensis*), Leatherback turtles (*Dermochelys coriacea*), and Banteng (*Bos javanicus*). Most recently, the last male Sumatran rhino in Malaysia also died in May 2019 [4] due to extremely poor conservation efforts in the country. Additionally, this scenario is also listed by the International Union for Conservation of Nature (IUCN) which more than 27,000 species globally are threatened with extinction, including 40% of amphibians, 25% of mammals, 14% of birds and 21% others [5].

In Malaysia, a conventional approach such as camera trapping has been regularly used as part of the conservation efforts. However, the visual-based inspection does not magnify the confidence in species identification compared to DNA-based technology. As stated in the country's National Policy on Biological Diversity 2016-2025 precisely on Goal 3, DNA profile databases need to be developed to enhance intelligence-led investigations and improve detection of illegal trade of Malaysian wildlife. Furthermore, the weaknesses in capacity management, lack of formal training for Next Generation Sequencing (NGS) laboratory work and bioinformatic analysis as well as shortage of funding [1] have also hampered the actions and efforts to reverse the alarming trends of species loss and restore the nation's biodiversity. Despite the catastrophic loss, Malaysia's biodiversity data is

still poorly characterized and systematically documented. This knowledge gap has been identified as one of the key factors impeding biodiversity monitoring efforts in Malaysia. Currently, most species data have limited visibility with very abysmal publications which are mostly restricted to morphological traits. To address this, the generation of genomic data could help revolutionize the understanding in biology and evolution of species thus enhance the intelligence-led investigations for the illegal trade of Malaysian wildlife.

Therefore, the mitochondrial genome (mitogenome) in the present study (summarized in Figure 1.1) attempts to generate an established reference mitogenome dataset for Malaysian fauna using the cutting-edge Next Generation Sequencing (NGS) technology coupled with High-Performance Computing (HPC). Mitogenomes are often sequenced especially for animals due to its small sizes and highly conserved [6]. Hence in future, the assembly of complete mitogenomes will be useful to understand the evolutionary relationships among taxa [7]. In developed countries such as USA, United Kingdom, Denmark, and Singapore, the rapid technological advancement of NGS has outperformed traditional Sanger sequencing technology by paving the way with a vast pool of genetic data at cost effective prices [29]. However, the current NGS-based research in Malaysia is moving at a slow pace because the NGS technology in the country is still considered high-priced. This is due to only abundance of prokaryotes has been sequenced compared to the complex eukaryote genomes. In order to accomplish the goal of this project, UTM has embarked on a collaboration with the Centre for GeoGenetics, University of Copenhagen, Denmark (UCPH) to assist with NGS facilities. With a state-of-art infrastructure, the centre provides formal NGS training in the laboratory and bioinformatic analysis for generating and analysing not only mitogenomes but also other massive omics projects.

This project will act as proof of concept where global collaboration effort could help expand our national capacity to promote conservation efforts and sustainability of the ecosystem in Malaysia. Furthermore, this significant effort of establishing DNA reference datasets using high-tech NGS technology will further magnifies the confidence in monitoring fauna diversity using invertebrate-derived

DNA (iDNA) [8–10], and expedite authorities to fight illegal smuggling of endangered species [11]. In addition, this project is in line with the country's National Policy on Biological Diversity 2016-2025, specifically Goal 3 and Goal 5 which is to establish a comprehensive DNA profile databases and thus ensure that all the active conservationist nations have adequate resources to effectively manage and monitor biodiversity in Malaysia. This project will also contribute to the global effort to sequence all life by 2028; Earth BioGenome Project 2018-2028 [12].

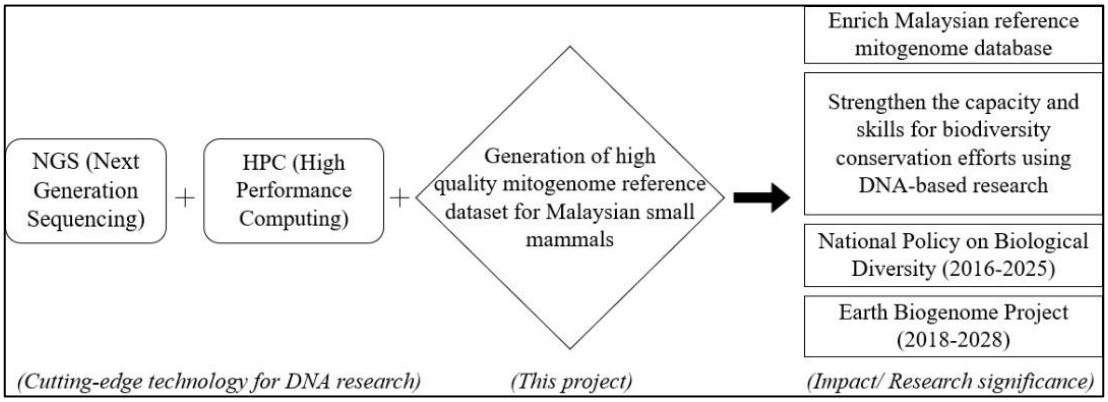

Figure 1.1 Overview of this research project and its significance impact

1.2 Problem statement

Despite the rich biodiversity in Malaysia, the most current dataset for eukaryote is limited to morphological data and short DNA barcodes. The situation is worsened by the lack of capacity and high technical skills required to generate and analyse complex genomic data. These challenges have remained a major obstacle for a large scale genomics sequencing effort thus hindering a more sophisticated data management and analysis and now de-rigueur in most active conservationist nations. With the advancement of Next Generation Sequencing (NGS) technology, generation of high quality mitogenome data offers a more promising approach and wider downstream application especially in biodiversity monitoring using DNA-based technology and wildlife forensics as well as species evolution and population genomics.

1.3 Objectives of the Study

(1) To generate high quality library from fresh Malaysian small mammal samples for shotgun sequencing

(2) To assemble high quality mitogenomes from shotgun data using a bioinformatic pipeline

(3) To annotate the assembled mitogenome according to the respective genes

1.4 Scope of the Study

This study generated Malaysian *Balionycteris maculata* (M1) and *Callosciurus notatus* (M8) mitogenomes reference dataset. Samples were obtained from Forest Research Institute Malaysia (FRIM) and were extracted in Universiti Teknologi Malaysia (UTM). Later, the library construction were conducted in Centre for GeoGenetics, University of Copenhagen (UCPH) prior to shotgun sequencing. The libraries were constructed using sheared DNA samples and rebuilt into Illumina-compatible sequencing (NGS) libraries using BEST protocol [13]. Then, the libraries were indexed and pooled before sequenced across a lane of BGISEQ-500. By connecting to HoloGenomics Servers located in Danish Center for Scientific Computing at Copenhagen University, the raw reads produced were trimmed and assembled using AdapterRemoval [14], MITOBIM [15] and PALEOMIX [68]. The newly assembled mitogenomes were generated with and without available reference genome in Genbank. Next, the mitogenomes were further annotated and will be deposited to Genbank for public use.

1.5 Significance of the Study

The work presented in this project focus on the generation of mitogenomes dataset for Malaysian small mammals. Apart from enriching the current reference database, the success of this project will ultimately strengthen the capacity and skills for biodiversity conservation efforts especially using DNA-based research. By using

the advanced NGS technology coupled with high-performance computing, a large-scale discovery of genetic markers could be generated and further translated to study evolutionary biology, population genetics, phylogeography, systematics and conservation. In addition, the downstream application of this work is valuable in advanced identification tools for wildlife forensics to monitor illegal trade of endangered species in the region. Besides, it will aid the future work on employing invertebrate-derived DNA (iDNA) for biodiversity monitoring. Furthermore, this work is one of the goals highlighted in the National Policy on Biological Diversity 2016-2025 [1] and in line with the global effort; Earth BioGenome Project 2018-2028 to sequence all life by 2028 [12]. The socioeconomic and ecology impact of this project if done in large-scale, would be beyond typical economic gains derived from taxonomic work applied in conventional biodiversity conservation management.

REFERENCES

1. *National Policy on Biological Diversity 2016 – 2025*. Ministry of Natural Resources and Environment (NRE), Malaysia. 2016.
2. Crist E, Mora C, Engelman R. The interaction of human population, food production, and biodiversity protection. *Science*. 2017. 356(6335):260–4.
3. Gristwood A. CSI wildlife: How science is helping detectives to track down poaching and illegal trade in wildlife. *EMBO Rep*. 2019. 20(1).
4. Bittel Jason (2019, May 27) Last male Sumatran rhino in Malaysia dies. *National Geographic*.
5. IUCN. The IUCN Red List of Threatened Species. Version 2019-1. 2019.
6. Land M, Hauser L, Jun S-R, Nookaew I, Leuze MR, Ahn T-H, et al. Insights from 20 years of bacterial genome sequencing. *Funct Integr Genomics*. 2015. 15(2):141–61.
7. Smith DR. The past, present and future of mitochondrial genomics: have we sequenced enough mtDNAs? *Brief Funct Genomics*. 2016. 15(1):47–54.
8. Drinkwater R, Schnell IB, Bohmann K, Bernard H, Veron G, Clare E, et al. Using metabarcoding to compare the suitability of two blood-feeding leech species for sampling mammalian diversity in North Borneo. *Mol Ecol Resour*. 2019. 19(1):105–17.
9. Schnell IB, Sollmann R, Calvignac-Spencer S, Siddall ME, Yu DW, Wilting A, et al. iDNA from terrestrial haematophagous leeches as a wildlife surveying and monitoring tool - prospects, pitfalls and avenues to be developed. *Front Zool*. 2015. 12:24.
10. Weiskopf SR, McCarthy KP, Tessler M. Using terrestrial haematophagous leeches to enhance tropical biodiversity monitoring programmes in Bangladesh. *Journal of applied*. 2018.
11. Børsting C, Morling N. Next generation sequencing and its applications in forensic genetics. *Forensic Sci Int Genet*. 2015. 18:78–89.
12. Lewin HA, Robinson GE, Kress WJ, Baker WJ, Coddington J, Crandall KA, et al. Earth BioGenome Project: Sequencing life for the future of life. *Proc Natl Acad Sci U S A*. 2018. 115(17):4325–33.

13. Carøe C, Gopalakrishnan S, Vinner L, Mak SST, Sinding MHS, Samaniego JA, et al. Single-tube library preparation for degraded DNA. *Methods Ecol Evol*. 2017. 9(2):410–9.
14. Schubert M, Lindgreen S, Orlando L. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Res Notes*. 2016. 9:88.
15. Hahn C, Bachmann L, Chevreux B. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. *Nucleic Acids Res*. 2013. 41(13):e129–e129.
16. World Wide Fund for Nature (WWF). WWF-Malaysia Annual Review. Malaysia. 2018.
17. Hebert PDN, Cywinska A, Ball SL, deWaard JR. Biological identifications through DNA barcodes. *Proc Biol Sci*. 2003. 270(1512):313–21.
18. Mukherjee S, Stamatis D, Bertsch J, Ovchinnikova G, Katta HY, Mojica A, et al. Genomes OnLine database (GOLD) v.7: updates and new features. *Nucleic Acids Res*. 2019. 47(D1):D649–59.
19. WWF. Living Planet Report - 2018: Aiming Higher. 2018.
20. Robideau GP, De Cock AWAM, Coffey MD, Voglmayr H, Brouwer H, Bala K, et al. DNA barcoding of oomycetes with cytochrome c oxidase subunit I and internal transcribed spacer. *Mol Ecol Resour*. 2011. 11(6):1002–11.
21. Koepfli K-P, Paten B, Genome 10K Community of Scientists, O'Brien SJ. The Genome 10K Project: a way forward. *Annu Rev Anim Biosci*. 2015. 3:57–111.
22. i5K Consortium. The i5K Initiative: advancing arthropod genomics for knowledge, human health, agriculture, and the environment. *J Hered*. 2013. 104(5):595–600.
23. GIGA Community of Scientists, Bracken-Grissom H, Collins AG, Collins T, Crandall K, Distel D, et al. The Global Invertebrate Genomics Alliance (GIGA): developing community resources to study diverse invertebrate genomes. *J Hered*. 2014. 105(1):1–18.
24. Zhang G, Rahbek C, Graves GR, Lei F, Jarvis ED, Gilbert MTP. Genomics: Bird sequencing project takes off. *Nature*. 2015. 522(7554):34.
25. Teeling EC, Vernes SC, Dávalos LM, Ray DA, Gilbert MTP, Myers E, et al. Bat Biology, Genomes, and the Bat1K Project: To Generate Chromosome-Level Genomes for All Living Bat Species. *Annu Rev Anim Biosci*. 2018. 6:23–46.
26. Bakhtiar MF, Too C-L, Tang M-M, Sulaiman S, Tan L-K, Ahmad-Fauzi NA, et

al. Incidental pharmacogenetics findings in an HLA-related research: Considerations for primary prevention. *Clin Exp Allergy*. 2019. 49(4):537–40.

27. Abou Diwan E, Zeitoun RI, Abou Haidar L, Cascorbi I, Khoueiry Zgheib N. Implementation and obstacles of pharmacogenetics in clinical practice: An international survey. *Br J Clin Pharmacol*. 2019.
28. Sims D, Sudbery I, Ilott NE, Heger A, Ponting CP. Sequencing depth and coverage: key considerations in genomic analyses. *Nat Rev Genet*. 2014. 15(2):121–32.
29. Helmy M, Awad M, Mosa KA. Limited resources of genome sequencing in developing countries: Challenges and solutions. *Appl Transl Genom*. 2016. 9:15–9.
30. Hodgkin J. Genome Size. In: Brenner S, Miller JH, editors. Encyclopedia of Genetics. *New York: Academic Press*. 2001. p. 865.
31. Sakaff MKLM, Rahman AYA, Saito JA, Hou S, Alam M. Complete genome sequence of the thermophilic bacterium Geobacillus thermoleovorans CCB_US3_UF5. *J Bacteriol*. 2012. 194(5):1239–1239.
32. Plieskatt JL, Rinaldi G, Feng Y, Levine PH, Easley S, Martinez E, et al. Methods and matrices: approaches to identifying miRNAs for nasopharyngeal carcinoma. *J Transl Med*. 2014. 6;12:3.
33. Hasmad HN, Lai KN, Wen WX, Park DJ, Nguyen-Dumont T, Kang PCE, et al. Evaluation of germline BRCA1 and BRCA2 mutations in a multi-ethnic Asian cohort of ovarian cancer patients. *Gynecol Oncol*. 2016. 141(2):318–22.
34. Teh LK, Subramaniam V, Tuan Abdu Aziz TA, Lee LS, Ismail MI, Yu CY, et al. Systematic characterization and comparison of the CYP2C9 variability of the Orang Asli in Malaysia with 12 populations. *Drug Metab Pharmacokinet*. 2016. 31(4):304–13.
35. Salleh MZ, Teh LK, Lee LS, Ismet RI, Patowary A, Joshi K, et al. Systematic pharmacogenomics analysis of a Malay whole genome: proof of concept for personalized medicine. *PLoS One*. 2013. 8(8):e71554.
36. Aarnink A, Apoil P-A, Takahashi I, Osada N, Blancher A. Characterization of MHC class I transcripts of a Malaysian cynomolgus macaque by high-throughput pyrosequencing and EST libraries. *Immunogenetics*. 2011. 63(11):703–13.
37. Cooper GM. The Cell: A Molecular Approach. *ASM Press*. 2000. 689 p.

38. Zhou X-H, Wan X-T, Jin Y-H, Zhang W. Concept of scientific wildlife conservation and its dissemination. *Zool Res*. 2016. 37(5):270–4.
39. Tucker MA, Rogers TL. Examining the prey mass of terrestrial and aquatic carnivorous mammals: minimum, maximum and range. *PLoS One*. 2014. 9(8):e106402.
40. Schalk CM, Trees T, Pierce JB. Food habits of sympatric pitvipers from the West Gulf Coastal Plain, USA. *Herpetological Review*. 2018. 49(1):1-5.
41. Masoero G, Morosinotto C, Laaksonen T, Korpimäki E. Food hoarding of an avian predator: sex- and age-related differences under fluctuating food conditions. *Behavioral Ecology and Sociobiology*. 2018. 72(10):159.
42. Goldingay RL, Carthew SM, Whelan RJ. The Importance of Non-Flying Mammals in Pollination. *Oikos*. 1991. 61(1):79–87.
43. Cárdenas S, Nivelo-Villavicencio C, Cárdenas JD, P OL, Tinoco BA. First record of flower visitation by a rodent in Neotropical Proteaceae, Oreocallis grandiflora. *J Trop Ecol*. 2017. 33(2):174–7.
44. Stein K, Coulibaly D, Stenchly K, Goetze D, Porembski S, Lindner A, et al. Bee pollination increases yield quantity and quality of cash crops in Burkina Faso, West Africa. *Sci Rep*. 2017. 7(1):17691.
45. Hooper DU, Adair EC, Cardinale BJ, Byrnes JEK, Hungate BA, Matulich KL, et al. A global synthesis reveals biodiversity loss as a major driver of ecosystem change. *Nature*. 2012. 486(7401):105–8.
46. Isbell F, Reich PB, Tilman D, Hobbie SE, Polasky S, Binder S. Nutrient enrichment, biodiversity loss, and consequent declines in ecosystem productivity. *Proc Natl Acad Sci U S A*. 2013. 110(29):11911–6.
47. Francis CM, Barrett P. *A Field Guide to the Mammals of South-East Asia*. 2nd ed. New Holland Publishers. 2008.
48. Teletchea F, Bernillon J, Duffraisse M. Molecular identification of vertebrate species by oligonucleotide microarray in food and forensic samples. *Journal of Applied*. 2008.
49. Everett KR, Rees-George J, Pushparajah I, Janssen BJ, Luo Z, Others. Advantages and disadvantages of microarrays to study microbial population dynamics- a minireview. *N Z Plant Prot*. 2010. 63:1–6.
50. Chin Chin T, Adibah AB, Danial Hariz ZA, Siti Azizah MN. Detection of mislabelled seafood products in Malaysia by DNA barcoding: Improving

transparency in food market. *Food Control*. 2016. 64:247–56.

51. El Sheikha AF, Mokhtar NFK, Amie C, Lamasudin DU, Isa NM, Mustafa S. Authentication technologies using DNA-based approaches for meats and halal meats determination. *Food Biotechnol*. 2017. 31(4):281–315.
52. Chao Z, Liao J, Liang Z, Huang S, Zhang L, Li J. Cytochrome C oxidase subunit I barcodes provide an efficient tool for Jinqian Baihua She (Bungarus parvus) authentication. *Pharmacogn Mag*. 2014. 10(40):449–57.
53. Kumar US, Ratheesh RV, Thomas G, George S. Use of DNA barcoding in wildlife forensics: a study of sambar deer (Rusa unicolor). *Forest Sci Technol*. 2012. 8(4):224–6.
54. Lo Y-T, Shaw P-C. DNA-based techniques for authentication of processed food and food supplements. *Food Chem*. 2018. 240:767–74.
55. Shen Y-Y, Chen X, Murphy RW. Assessing DNA barcoding as a tool for species identification and data quality control. *PLoS One*. 2013. 8(2):e57125.
56. Bruford MW, Davies N, Dulloo ME, Faith DP, Walters M. Monitoring Changes in Genetic Diversity. *The GEO Handbook on Biodiversity Observation Networks*. 2017.
57. Frankham R. Genetic rescue of small inbred populations: Meta-analysis reveals large and consistent benefits of gene flow. *Mol Ecol*. 2015. 24(11):2610–8.
58. Boore JL. Animal mitochondrial genomes. *Nucleic Acids Res*. 1999. 27(8):1767–80.
59. Taanman JW. The mitochondrial genome: structure, transcription, translation and replication. *Biochim Biophys Acta*. 1999. 1410(2):103–23.
60. Hebert PDN, Ratnasingham S, deWaard JR. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proc Biol Sci*. 2003. 270 Suppl 1:S96–9.
61. Kartavtsev YP, Park T-J, Vinnikov KA, Ivankov VN, Sharina SN, Lee J-S. Cytochrome b (Cyt-b) gene sequence analysis in six flatfish species (Teleostei, Pleuronectidae), with phylogenetic and taxonomic insights. *Mar Biol*. 2007. 152(4):757–73.
62. Vences M, Thomas M, van der Meijden A, Chiari Y, Vieites DR. Comparative performance of the 16S rRNA gene in DNA barcoding of amphibians. *Front Zool*. 2005. 16;2(1):5.
63. Vieira GA, Prosdocimi F. Accessible molecular phylogenomics at no cost:

obtaining 14 new mitogenomes for the ant subfamily Pseudomyrmecinae from public data. *PeerJ*. 2019. 7:e6271.

64. Winkelmann I, Campos PF, Strugnell J, Cherel Y, Smith PJ, Kubodera T, et al. Mitochondrial genome diversity and population structure of the giant squid Architeuthis: genetics sheds new light on one of the most enigmatic marine species. *Proc Biol Sci*. 2013. 280(1759):20130273.
65. Schnell IB, Thomsen PF, Wilkinson N, Rasmussen M, Jensen LRD, Willerslev E, et al. Screening mammal biodiversity using DNA from leeches. *Curr Biol*. 2012. 22(8):R262–3.
66. Ghorbani A, Saeedi Y, de Boer HJ. Unidentifiable by morphology: DNA barcoding of plant material in local markets in Iran. *PLoS One*. 2017. 12(4):e0175722.
67. Head SR, Komori HK, LaMere SA, Whisenant T, Van Nieuwerburgh F, Salomon DR, et al. Library construction for next-generation sequencing: overviews and challenges. *Biotechniques*. 2014. 56(2):61–4, 66, 68.
68. de Freitas PD, Mendez FL, Chávez-Congrains K, Galetti PM Jr, Coutinho LL, Pissinatti A, et al. Next-Generation Sequencing of the Complete Mitochondrial Genome of the Endangered Species Black Lion Tamarin Leontopithecus chrysopygus (Primates) and Mitogenomic Phylogeny Focusing on the Callitrichidae Family. *G3*. 2018. 8(6):1985–91.
69. O'Connor B, Secades C, Penner J, Sonnenschein R, Skidmore A, Burgess ND, et al. Earth observation as a tool for tracking progress towards the Aichi Biodiversity Targets. *Remote Sens Ecol Conserv*. 2015. 1(1):19–28.
70. Ariffin M. Recent Challenges for the Enforcement of Wildlife Laws in East Malaysia. *Pertanika Journal of Social Sciences & Humanities*. 2018. 183
71. Liu L, Li Y, Li S, Hu N, He Y, Pong R, et al. Comparison of next-generation sequencing systems. *J Biomed Biotechnol*. 2012. 2012:251364.
72. Mak SST, Gopalakrishnan S, Carøe C, Geng C, Liu S, Sinding M-HS, et al. Comparative performance of the BGISEQ-500 vs Illumina HiSeq2500 sequencing platforms for palaeogenomic sequencing. *Gigascience*. 2017. 6(8):1–13.
73. Poptsova MS, Il'icheva IA, Nechipurenko DY, Panchenko LA, Khodikov MV, Oparina NY, et al. Non-random DNA fragmentation in next-generation sequencing. *Sci Rep*. 2014. 4:4532.

74. Thorstenson YR, Hunicke-Smith SP, Oefner PJ, Davis RW. An automated hydrodynamic process for controlled, unbiased DNA shearing. *Genome Res*. 1998. 8(8):848–55.
75. Aigrain L, Gu Y, Quail MA. Quantitation of next generation sequencing library preparation protocol efficiencies using droplet digital PCR assays - a systematic comparison of DNA library preparation kits for Illumina sequencing. *BMC Genomics*. 2016. 17:458.
76. Robin JD, Ludlow AT, LaRanger R, Wright WE, Shay JW. Comparison of DNA Quantification Methods for Next Generation Sequencing. *Sci Rep*. 2016. 6:24067.
77. Fracassetti M, Griffin PC, Willi Y. Validation of Pooled Whole-Genome Re-Sequencing in Arabidopsis lyrata. *PLoS One*. 2015. 10(10):e0140462.
78. Masetti R, Castelli I, Astolfi A, Bertuccio SN, Indio V, Togni M, et al. Genomic complexity and dynamics of clonal evolution in childhood acute myeloid leukemia studied with whole-exome sequencing. *Oncotarget*. 2016. 7(35):56746–57.
79. Dierckxsens N, Mardulyn P, Smits G. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucleic Acids Res*. 2017. 45(4):e18.
80. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol*. 2012. 19(5):455–77.
81. Andrews S, Others. FastQC: a quality control tool for high throughput sequence data. 2016.
82. Schubert M, Ermini L, Der Sarkissian C, Jónsson H, Ginolhac A, Schaefer R, et al. Characterization of ancient and modern genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX. *Nat Protoc*. 2014. 9(5):1056–82.
83. Wysoker A, Tibbetts K, Fennell T. Picard tools version 1.90. 2013;107(17):308.
84. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*. 2009. 25(14):1754–60.
85. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010. 20(9):1297–303.

86. Ginolhac A, Rasmussen M, Gilbert MTP, Willerslev E, Orlando L. mapDamage: testing for damage patterns in ancient DNA sequences. *Bioinformatics*. 2011. 27(15):2153–5.
87. Li H, Gao J, Liu H, Liu H, Liang A, Zhou X, et al. The architecture and complete sequence of mitochondrial genome of an assassin bug Agriosphodrus dohrni (Hemiptera: Reduviidae). *Int J Biol Sci*. 2011. 7(6):792–804.
88. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics*. 2009. 25(9):1189–91.
89. Milne I, Stephen G, Bayer M, Cock PJA, Pritchard L, Cardle L, et al. Using Tablet for visual exploration of second-generation sequencing data. *Brief Bioinform*. 2013. 14(2):193–202.
90. Dominguez Del Angel V, Hjerde E, Sterck L, Capella-Gutierrez S, Notredame C, Vinnere Pettersson O, et al. Ten steps to get started in Genome Assembly and Annotation. *F1000Res*. 2018. 7.
91. Wyman SK, Jansen RK, Boore JL. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics*. 2004. 20(17):3252–5.
92. Bernt M, Donath A, Jühling F, Externbrink F, Florentz C, Fritzsch G, et al. MITOS: improved de novo metazoan mitochondrial genome annotation. *Mol Phylogenet Evol*. 2013. 69(2):313–9.
93. Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, et al. GeSeq - versatile and accurate annotation of organelle genomes. *Nucleic Acids Res*. 2017. 45(W1):W6–11.
94. Meng G, Li Y, Yang C, Liu S. MitoZ: a toolkit for animal mitochondrial genome assembly, annotation and visualization. *Nucleic Acids Res*. 2019. 47(11):e63.
95. Florea L, Souvorov A, Kalbfleisch TS, Salzberg SL. Genome assembly has a major impact on gene content: a comparison of annotation in two Bos taurus assemblies. *PLoS One*. 2011. 6(6):e21400.
96. Sayadi A, Immonen E, Tellgren-Roth C, Arnqvist G. The Evolution of Dark Matter in the Mitogenome of Seed Beetles. *Genome Biol Evol*. 2017. 9(10):2697–706.
97. Salzberg SL. Next-generation genome annotation: we still struggle to get it right. *Genome Biol*. 2019. 20(1):92.
98. Simbolo M, Gottardi M, Corbo V, Fassan M, Mafficini A, Malpeli G, et al. DNA

qualification workflow for next generation sequencing of histopathological samples. *PLoS One*. 2013. 8(6):e62692.

99. O'Donnell JL, Kelly RP, Lowell NC, Port JA. Indexed PCR Primers Induce Template-Specific Bias in Large-Scale DNA Sequencing Studies. PLoS One. 2016. 11(3):e0148698.
100. Bohmann K, Evans A, Gilbert MTP, Carvalho GR, Creer S, Knapp M, et al. Environmental DNA for wildlife biology and biodiversity monitoring. *Trends Ecol Evol*. 2014. 29(6):358–67.
101. DeAngelis MM, Wang DG, Hawkins TL. Solid-phase reversible immobilization for the isolation of PCR products. *Nucleic Acids Res*. 1995. 23(22):4742–3.
102. Hawkins TL, O'Connor-Morin T, Roy A, Santillan C. DNA purification and isolation using a solid-phase. *Nucleic Acids Res*. 1994. 22(21):4543–4.
103. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. *BMC Bioinformatics*. 2009. 10:421.
104. Nguyen HQ, Kim YI, Borzée A, Jang Y. Efficient isolation method for high-quality genomic DNA from cicada exuviae. *Ecol Evol*. 2017. 7(20):8161–9.
105. Tin MM-Y, Economo EP, Mikheyev AS. Sequencing degraded DNA from non-destructively sampled museum specimens for RAD-tagging and low-coverage shotgun phylogenetics. *PLoS One*. 2014. 9(5):e96793.
106. Matsvay AD, Alborova IE, Pimkina EV, Markelov ML, Khafizov K, Mustafin KK. Experimental approaches for ancient DNA extraction and sample preparation for next generation sequencing in ultra-clean conditions. *Conserv Genet Resour*. 2018.1–9.
107. Knierim E, Lucke B, Schwarz JM, Schuelke M, Seelow D. Systematic comparison of three methods for fragmentation of long-range PCR products for next generation sequencing. *PLoS One*. 2011. 6(11):e28240.
108. Fernández-Soto P, Gandasegui Arahuetes J, Sánchez Hernández A, López Abán J, Vicente Santiago B, Muro A. A loop-mediated isothermal amplification (LAMP) assay for early detection of Schistosoma mansoni in stool samples: a diagnostic approach in a murine model. *PLoS Negl Trop Dis*. 2014. 8(9):e3126.
109. Sinha R, Stanley G, Gulati GS, Ezran C, Travaglini KJ. Index switching causes "spreading-of-signal" among multiplexed samples in Illumina HiSeq 4000 DNA sequencing. *BioRxiv*. 2017.
110. Psifidi A, Dovas CI, Bramis G, Lazou T, Russel CL, Arsenos G, et al.

Comparison of eleven methods for genomic DNA extraction suitable for large-scale whole-genome genotyping and long-term DNA banking using blood samples. *PLoS One*. 2015. 10(1):e0115960.

111. Bacich DJ, Sobek KM, Cummings JL, Atwood AA, O'Keefe DS. False negative results from using common PCR reagents. *BMC Res Notes*. 2011. 4:457.
112. Pan P, Jin W, Li X, Chen Y, Jiang J, Wan H, et al. Optimization of multiplex quantitative polymerase chain reaction based on response surface methodology and an artificial neural network-genetic algorithm approach. *PLoS One*. 2018. 13(7):e0200962.
113. Chatterjee N, Banerjee T, Datta S. Accurate estimation of nucleic acids by amplification efficiency dependent PCR. *PLoS One*. 2012. 7(8):e42063.
114. Witek MA, Llopis SD, Wheatley A, McCarley RL, Soper SA. Purification and preconcentration of genomic DNA from whole cell lysates using photoactivated polycarbonate (PPC) microfluidic chips. *Nucleic Acids Res*. 2006. 34(10):e74.
115. Fisher S, Barry A, Abreu J, Minie B, Nolan J, Delorey TM, et al. A scalable, fully automated process for construction of sequence-ready human exome targeted capture libraries. *Genome Biol*. 2011. 12(1):R1.
116. Anand S, Mangano E, Barizzone N, Bordoni R, Sorosina M, Clarelli F, et al. Next Generation Sequencing of Pooled Samples: Guideline for Variants' Filtering. *Sci Rep*. 2016. 6:33735.
117. Mohd Salleh F, Ramos-Madrigal J, Peñaloza F, Liu S, Mikkel-Holger SS, Riddhi PP, et al. An expanded mammal mitogenome dataset from Southeast Asia. *Gigascience*. 2017. 6(8):1–8.
118. Schbath S, Martin V, Zytnicki M, Fayolle J, Loux V, Gibrat J-F. Mapping reads on a genomic sequence: an algorithmic overview and a practical comparative analysis. *J Comput Biol*. 2012. 19(6):796–813.
119. Timbó RV, Togawa RC, Costa MMC, Andow DA, Paula DP. Mitogenome sequence accuracy using different elucidation methods. *PLoS One*. 2017. 12(6):e0179971.
120. Chhangawala S, Rudy G, Mason CE, Rosenfeld JA. The impact of read length on quantification of differentially expressed genes and splice junction detection. *Genome Biol*. 2015. 16:131.
121. Falkenberg M. Mitochondrial DNA replication in mammalian cells: overview of

the pathway. *Essays Biochem*. 2018. 62(3):287–96.

122. Poptsova MS, Gogarten JP. Using comparative genome analysis to identify problems in annotated microbial genomes. *Microbiology*. 2010. 156(Pt 7):1909–17.

123. Hooper DU, Chapin FS, Ewel JJ, Hector A. Effects of biodiversity on ecosystem functioning: a consensus of current knowledge. *Ecological*. 2005.

124. Tropical Rainforests Activity. New England Primate Conservancy.

125. Field D, Davies N. Biocode: The New Age of Genomics. *Oxford University Press.* 2015. 196 p.

126. Díaz S, Purvis A, Cornelissen JHC, Mace GM, Donoghue MJ, Ewers RM, et al. Functional traits, the phylogeny of function, and ecosystem service vulnerability. *Ecol Evol*. 2013. 3(9):2958–75.

127. Buchanan AV, Sholtis S, Richtsmeier J, Weiss KM. What are genes "for" or where are traits "from"? What is the question? *Bioessays*. 2009. 31(2):198–208.

128. Kakumanu R, Hodgson WC, Ravi R, Alagon A, Harris RJ, Brust A, et al. Vampire Venom: Vasodilatory Mechanisms of Vampire Bat (Desmodus rotundus) Blood Feeding. *Toxins*. 2019. 11(1).

129. Ma D, Mizurini DM, Assumpção TCF, Li Y, Qi Y, Kotsyfakis M, et al. Desmolaris, a novel factor XIa anticoagulant from the salivary gland of the vampire bat (Desmodus rotundus) inhibits inflammation and thrombosis in vivo. *Blood*. 2013. 122(25):4094–106.

130. Hodgkison R. The ecology of fruit bats (Chiroptera: Pteropodidae) in a Malaysian lowland dipterocarp forest, with particular reference to the spotted-winged fruit bat (Balionycteris maculata, Thomas). *University of Aberdeen.* 2001.

131. Willmer P. Pollination and Floral Ecology. *Princeton University Press*. 2011.

132. Douglass M. Animals you might not know pollinate flowers. *BBC*. 2015.

133. Wong S, Lau S, Woo P, Yuen K-Y. Bats as a continuing source of emerging infections in humans. *Reviews in Medical Virology*. 2007. 17:67–91.

134. Kasso M, Balakrishnan M. Ecological and economic importance of bats (Order Chiroptera). *ISRN Biodiversity*. 2013.

135. Wibbelt G, Kurth A, Yasmum N, Bannert M, Nagel S, Nitsche A, et al. Discovery of herpesviruses in bats. *J Gen Virol*. 2007. 88(Pt 10):2651–5.

136. Mandl JN, Schneider C, Schneider DS, Baker ML. Going to Bat(s) for Studies

of Disease Tolerance. *Front Immunol*. 2018. 9:2112.

137. Fenton MB, Davison M, Kunz TH, McCracken GF, Racey PA, Tuttle MD. Linking bats to emerging diseases. *Science*. 2006. 311(5764):1098–9

138. Haake DA, Levett PN. Leptospirosis in humans. *Curr Top Microbiol Immunol*. 2015. 387:65–97.

139. Drexler JF, Corman VM, Müller MA, Maganga GD, Vallo P, Binger T, et al. Bats host major mammalian paramyxoviruses. *Nat Commun*. 2012. 3:796.

140. Sarmashghi S, Bohmann K, Gilbert MTP, Bafna V. Assembly-free and alignment-free sample identification using genome skims. *bioRxiv*. 2018.

LIST OF PUBLICATIONS

1. Puteri Nur Syahzanani Jahari, Mohd Shahir Shamsir, Faezah Mohd Salleh. Monash Science Symposium. 2018. PRISMA review: Current State of Genomic Sequencing in Malaysia. Pg 156-159. Monash Science University. ISBN 978-967-5227-30-1.
2. Bashir Mohammed Abubakar, Puteri Nur Syahzanani Jahari, Ooi Zhi Sin, Thong Hui Yee, Mohd Shahir Shamir Omar, Alina Wagiran and Faezah Mohd Salleh. Evaluation of the Ficus deltoidea (Mas Cotek) Herbal Medicinal Products (HMPs) Authenticity via DNA Barcoding and HPLC Analysis (under review)
3. Ooi Zhi Sin, Puteri Nur Syahzanani Jahari, Kah Shean Sim, Shi Xiang Foo, Nurain Najwa Mohd Zawai and Faezah Mohd Salleh. DNA Barcoding of Commercial Fish Products using Dual Mitochondrial Marker Exposes Evidence for Mislabeling and Trade of Endangered Species (under review)