GENOMIC ANALYSES AND CHARACTERIZATION ON THE TEMPORAL AND QUANTITATIVE FATES OF HUMAN DNA AFTER BLOOD MEALS OF TROPICAL BED BUG, *Cimex hemipterus* (F.)

LIM LI

UNIVERSITI SAINS MALAYSIA

2023

GENOMIC ANALYSES AND CHARACTERIZATION ON THE TEMPORAL AND QUANTITATIVE FATES OF HUMAN DNA AFTER BLOOD MEALS OF TROPICAL BED BUG, *Cimex hemipterus* (F.)

by

LIM LI

Thesis submitted in fulfilment of the requirements
for the degree of
Doctor of Philosophy

June 2023

ACKNOWLEDGEMENT

I am grateful to those who helped me complete this research, and I would like to divulge their contributions individually.

First, I sincerely appreciate my supervisor, Associate Professor Ts. Dr Abdul Hafiz Ab Majid, for his great assistance and guidance during my research, especially during the COVID pandemic. He always ensured I acquired any aid or materials needed to run my experiment. So I could complete my research on time. Moreover, he also gave lots of precious suggestions for developing this thesis. My appreciation also extends to the Ministry of Education (MOE), as this study was funded by the Fundamental Research Grant (FRGS) (203/PBIOLOGI/6711681) supported by MOE, Malaysia.

Next, I would like to credit my senior, Mr Abd Hafis Abd Rahim, who had given me lots of assistance in sample collection. He was also the volunteer who managed the human blood feed bed bugs to maintain the laboratory colony for experiment purposes and also offered the blood-fed bed bugs samples for my research. I might also add, parenthetically, the Universiti Sains Malaysia, Human Ethics Committee (USM/JEPeM/19120868) had approved this protocol of human blood feed for bed bugs colony maintenance. As for my laboratory colleague, Mr Lim Li Yang, I would also like to thank him for giving me a hand in my bed bugs samples collection.

Last but not least, I am highly obliged to my father, Lim Soon Leong, my mother, Ho Piek Har, my sister, Lim Min, and my brothers, Lim Chee Khee and Lim Chee Wah, for their continuous encouragement and unfailing support throughout the research process and for completing this thesis.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

LIST OF PLATES

Page

LIST OF SYMBOLS

%	Percentage
&	And
®	Registered
µℓ	Microlitre
µm	Micrometre
g	Gram
ℓ	Litre
mℓ	Millilitre
mm	Millimetre
™	Trademark

LIST OF ABBREVIATIONS

FLASH	Fast Length Adjustment of Short Reads
SPAdes	St. Petersburg genome assembler
QUAST	Quality Assessment Tool
BUSCO	Benchmarking Universal Single-Copy Orthologues
MC1R	Melanocortin 1 receptor
OCA2	Oculocutaneous albinism II
HERC2	HECT and RLD domain containing E3 ubiquitin protein ligase 2
ASIP	Agouti Signalling Protein
IRF4	Interferon regulatory factor 4
TYR	Tyrosinase
SLC24A4	Solute Carrier Family 24 member 4
SLC24A5	Solute Carrier Family 24 member 5
RPL18	60S ribosomal protein L18
PRSS1	Trypsin-1
Ppk28	Pickpocket protein 28
Oacyl	O-acyltransferase like protein
LIM	Linl-1, Isl-1 and Mec-3

ANALISIS GENOMIK DAN PERINCIAN KE ATAS NASIB TEMPORAL DAN KUANTITATIF DNA MANUSIA SELEPAS SAJIAN DARAH PEPIJAT TROPIKA, *Cimex hemipterus* (F.)

ABSTRAK

Penyelidikan ini mengkaji potensi *Cimex hemipterus* (F.) sebagai serangga pemakan darah dalam bidang forensik dan juga mengkaji genom draf, metagenom, dan transkriptom spesies ini. Pertama-tama, genom draf *C. hemipterus* telah dikembangkan menggunakan perakitan *de novo*, dan genom draf ini digunakan sebagai rujukan dalam kajian lanjutan. Dalam kajian berkaitan forensik, kemungkinan untuk mengesan DNA manusia dalam *C. hemipterus* jantan dan betina selepas makan darah pada beberapa masa (Hari 0, 1, 3, 5, 7, 14, 30, dan 45) telah dikaji menggunakan DNA nuklear (nDNA) dan DNA mitokondria (mtDNA). Hasil kajian menunjukkan bahawa kedua-dua nDNA manusia dan mtDNA dapat dikesan dalam *C. hemipterus* jantan dan betina selepas makan darah dan dikesan sehingga 45 hari tanpa variasi yang signifikan dalam kepekatan DNA dan keberjayaan pengkayaan antara jantina. Berdasarkan kejayaan pengesanan DNA, penyelidikan ini kemudiannya meneruskan profil pengulangan pendek (STR) dan polimorfisme nukleotida tunggal (SNP) menggunakan hanya *C. hemipterus* jantan. Untuk STR, sampel DNA pada hari ke-0, ke-5, ke-14, ke-30, dan ke-45 telah digunakan dengan menggunakan 13 penanda STR. Penanda yang sama juga telah digunakan pada sampel *C. hemipterus* yang dikutip di lapangan. Hasil kajian menunjukkan bahawa profil STR lengkap dapat diperoleh daripada sampel D0, dan profil separa dapat diperoleh sehingga hari ke-45 selepas makan darah. Bagi sampel di lapangan, profil campuran dapat dikesan,

yang menunjukkan potensi kaedah DNA forensik dalam mengenalpasti sumber manusia *C. hemipterus*. Dalam kes profil SNP, set sampel yang sama (tidak termasuk sampel yang dikutip di lapangan) yang digunakan untuk profil STR telah dianalisis menggunakan 41 penanda SNP. Sama seperti hasil profil STR, profil lengkap diperoleh daripada sampel hari ke-0, dan profil separa dihasilkan sehingga 45 hari selepas makan darah. Untuk metagenom, komuniti bakteria dalam *C. hemipterus* selepas makan darah pada hari ke-0 dan ke-45 telah dianalisis dan dicirikan dengan memperkuatkan rantau hipervariabel v3-v4 gen 16S rRNA. Hasil kajian menunjukkan bahawa komposisi bakteria tidak berbeza secara signifikan antara *C. hemipterus* yang telah makan darah dan yang tidak, dan semua sampel mempunyai unit taksonomi operasi (OTU) dominan yang sama, iaitu *Wolbachia*, diikuti oleh *Dickeya chrysanthemi* dan *Pseudomonas*. Walau bagaimanapun, kepadatan bakteria dominan ini lebih tinggi pada *C. hemipterus* yang tidak makan. Terakhir, genom draf yang sebelum ini dihasilkan digunakan sebagai rujukan untuk perakitan transkriptom. Transkriptom *C. hemipterus* kemudiannya digunakan sebagai rujukan dalam analisis perbezaan ekspresi gen antara *C. hemipterus* selepas makan darah pada hari ke-0, ke-5, dan ke-45. Analisis ini mengenalpasti gen yang diekspresikan secara berbeza (DEGs) yang berkaitan dengan pencernaan darah *C. hemipterus*, termasuk "Peptida kardiaktif," "Faktor ADP-ribosilasi 6," "Cathepsin L1-suka," "Putatif aminopeptidase-2," "Chymotrypsin-2-suka," "Serin protease seperti chymotrypsin," "Pepsin-2B-suka," "Lipase triasilgliserol pankreas-suka," "Protein prolin-kaya liur asas 2," "Probable aquaporin TIP1-2," dan "Submandibular glandular kallikrein-9-suka." DEGs yang berkaitan dengan regulasi, metabolisme, pengangkutan, gerakan, respons imun dan tindak balas tekanan, endositosis, dan transduksi isyarat juga telah dikenalpasti dan dibincangkan.

GENOMIC ANALYSES AND CHARACTERIZATION ON THE TEMPORAL AND QUANTITATIVE FATES OF HUMAN DNA AFTER BLOOD MEALS OF TROPICAL BED BUG, *Cimex hemipterus* (F.)

ABSTRACT

This research exploits the potential of *Cimex hemipterus* (F.) as a blood-feeding insect in forensics and also studies the draft genome, metagenome and transcriptome of this species. First of all, the draft genome of *C. hemipterus* was developed using *de novo* assembly, and this draft genome was used as a reference in a later study. In forensics-related studies, the feasibility of detecting human DNA in male and female *C. hemipterus* post-blood meal at several time points (Day 0, 1, 3, 5, 7, 14, 30 and 45) was first done using nuclear DNA (nDNA) and mitochondria DNA (mtDNA) and the result showing both human nDNA and mtDNA could be detected from both male and female *C. hemipterus* post-blood meal and detected up to 45 days with no significant variations in DNA concentrations and amplification success between the sexes. Building upon the successful DNA detection, the research proceeded to Short Tandem Repeats (STR) and Single Nucleotide Polymorphism (SNP) profiling using only male bed bugs. For STR, DNA samples at day-0, -5, -14, -30, and -45 were employed using 13 STR markers. The same markers were also applied on field-collected *C. hemipterus*. The result showed that complete STR profile could be obtained from the D0 sample, and partial profiles could be obtained until day 45 post-blood meal. As for field samples, mix profiles could be detected, demonstrating the potential of forensic DNA methods in identifying the human DNA of *C. hemipterus*. In the case of SNP profiling, the same set of samples (excluding field-collected samples) used for STR profiling was analyzed using 41 SNP markers with

41 SNP markers. Similar to the STR profiling results, complete profiles were obtained from the day 0 sample, and partial profiles were generated up to 45 days post-blood meal. For metagenome, the bacteria communities in the D0 and D45 post-feeding *C. hemipterus* were analyzed and characterized by amplifying the v3-v4 hypervariable region of the 16S rRNA gene region. Results showed that the bacterial compositions have no significant difference between blood-fed and starved *C. hemipterus*, and all the samples have the same dominant OTUs, which is the *Wolbachia*, followed by *Dickeya chrysanthemi* and *Pseudomonas*. However, the densities of these dominant bacteria were higher in starved bed bugs. Last, the previously generated draft genome was used as a reference for transcriptome assembly. The transcriptome of *C. hemipterus* was then used as a reference in differential expression analysis between D0, D5 and D45 post-feeding *C. hemipterus*. This analysis revealed the differentially expressed genes (DEGs) that were probably associated with *C. hemipterus* blood digestion which included "Cardioactive peptide", "ADP-ribosylation factor 6", "Cathepsin L1-like", "Putative aminopeptidase-2", "Chymotrypsin-2-like", "Chymotrypsin-like serine proteinase", "Pepsin-2B-like", "Pancreatic triacylglycerol lipase-like", "Basic salivary proline-rich protein 2", "Probable aquaporin TIP1-2", and "Submandibular glandular kallikrein-9-like". DEGs related to regulation, metabolism, transport, motility, immune and stress response, endocytosis, and signal transduction were also identified and discussed.

CHAPTER 1

GENERAL INTRODUCTION

Evidence containing DNA at a crime scene can help link an individual to a crime. Common substrates from which DNA can be obtained at a crime scene include different types of fabric, objects used to drink or eat, or body fluids deposited on clothing (Gray et al., 2020). However, there are crime scenes where no noticeable biological stains can be found. In this case, other sources, such as insects found on the site and commonly consume human biological fluid, could be considered (Vieira et al., 2017).

Indeed, several studies have shown that DNA recovered from insect sources could be amplified, analysed, and used to identify the individual's identity. According to Gokool et al. (1993), human DNA could be detected from mosquito blood meals up to 26 hours after ingestion. Durdle et al. (2013) also reported that human DNA had been successfully recovered from blowfly artefacts after human blood, semen, and saliva consumption. Another study by Spitaleri et al. (2006) reported that the victim of murder casework in Sicily was successfully identified based on the generated DNA profile recovered from a smashed mosquito found on the crime scene or primary suspect's residence.

Although not typically considered forensically relevant substrates, tropical bed bugs, *Cimex hemipterus*, as an important urban pest in Malaysia (Zahran et al., 2016), can be a potential source of human DNA at a crime scene as these ectoparasites feed upon and digest human blood (Benoit et al., 2016). In a hypothetical scenario, bed bugs could feed on either victim or the suspect and stay captured at the enclosed crime scene, such as a room or a vehicle. Thus, a post-feeding bed bug might be a link between the suspect and the crime scene or a victim and the crime scene if the cadaver cannot be found (Curic

et al., 2014). In fact, Szalanski et al. (2006) performed the isolation and amplification of human DNA from the common bed bug, *Cimex lectularius*, by utilising the forensic STR marker D18S51.

Nevertheless, there is no information regarding human DNA being isolated and amplified from tropical bed bugs, *C. hemipterus*. Thus, one of the justifications of this study is to generate DNA profiles through the genetic material present in tropical bed bugs. The preliminary step would be exposing a group of laboratory *C. hemipterus* to feed with a volunteer's blood and determine the time interval of possibly detecting traces of the human DNA after bed bugs' blood meal.

In forensic science, human identification based on individuals' DNA profiling is the criminal case's primary goal (Zahra et al., 2018). Nowadays, Short Tandem Repeats (STR) are the most analysed genetic polymorphism in forensic genetics since it has introduced into casework in the mid-1990s because STR could provide robust analysis as the loci are highly sensitive and discriminatory (Goodwin et al., 2011). Besides STR analysis, single nucleotide polymorphism (SNP) analysis is an alternative identity profiling method. The advantage of using the SNP marker is that its amplicon length can be designed as short as possible to achieve a high success rate for degraded samples in individual identification. Moreover, some SNP markers can predict an individual's biogeographical ancestry and phenotypes, such as hair, eye, and skin colours, which can provide additional investigation leads (Vitoševic et al., 2019). Therefore, if full STR and SNP profiles could be generated from human DNA acquired from tropical bed bugs, the application of tropical bed bugs in forensics could be established.

Many insects establish symbiotic interactions with microorganisms, especially bacteria, that are essential in their host's life (Castillo et al., 2020). Thus, the digestion of

blood that contains human DNA in bed bugs could depend on the insect's digestive enzymes and bacterial activity. The duration and manner in which bacteria require to digest the blood meal might impact the possibility of detecting human DNA in tropical bed bugs, thus influencing its potential applications in the field of forensics. However, the bacteria communities that present within *C. hemipterus* have remained poorly characterised, except for a constantly isolated bacterium, *Wolbachia*, which is well-known as a mutualist of bed bugs that is important in biosynthesises of biotin and riboflavin for its nutritionally deficient host (Hosokawa et al., 2010; Evison et al., 2018; Fisher et al., 2019). Before further on the functions of each bacterium present in bed bugs, the first step is to characterise bacteria species present within the tropical bed bugs. With high-throughput sequencing technology development, the bacteria composition and their possible functional characteristics could be explored through sequencing without culturing the bacteria (Maccaferri et al., 2011). Therefore, the bacteria composition of blood-fed and starved *C. hemipterus* will be characterised through 16S rRNA metabarcoding to compare the community structure and deduce the possible bacteria species involved in blood meal digestions.

Except for bacteria symbionts, the detailed mechanisms of blood meal degradation within the digestive system of tropical bed bugs are still largely elusive. In order to provide molecular insights or identify the critical transcripts that are activated or suppressed during the blood meal processing, the laboratory strain of *C. hemipterus* will be used to study the transcriptome changes after feeding on blood meals through high-throughput RNA-sequencing.

Therefore, the objectives of the present study were:

1. To generate and develop draft genome assembly of *C. hemipterus*
2. To determine the differential detectability of partial sequence of human nuclear DNA (nDNA) and mitochondria DNA (mtDNA) in male and female *C. hemipterus* post-blood meal at several time points
3. To determine the feasibility of obtaining STR profile from laboratory-reared *C. hemipterus* post-blood meal at several time points and to determine the capability of this STR profiling to identify the human DNA of field-collected *C. hemipterus*
4. To determine the genotyping feasibility of SNP from the HIrisPlex-S system and its accuracy for phenotype prediction by utilizing DNA samples obtained from *C. hemipterus* post-blood meal at several time points
5. To determine and compare the bacteria compositions between days 0 and 45 post-feeding *C. hemipterus* through 16S rRNA metabarcoding
6. To identify the differentially expressed genes between days 0, 5 and 45 post-feeding *C. hemipterus* through transcriptome and differential expression analyses

CHAPTER 2

LITERATURE REVIEW

2.1 Forensic entomology

Forensic entomology is a branch of science in forensics that uses information about the lifecycles and behaviours of insects and other arthropods to help in criminal investigation and evidence interpretation in a legal context relating to humans and wildlife (Gennard, 2012).

Since the late 19th century, insects have been routinely used in forensic investigations. A decomposing body is an excellent food source and attracts a specialized insect community. However, different degrees of decay could attract different types of necrophilous insects. The difference in fondness creates a succession of insects on corpses, which occurs in a predictable sequence with a set of known times over the several stages of decay. Thus, these insects can be employed to estimate the post-mortem intervals (PMI) (Amendt et al., 2004; Joseph et al., 2011; Gennard, 2012). Other than insects' succession, the age and development of maggots could also be used to infer an approximate time frame of PMI (Volckaert, 2020).

Although forensic entomology always seems to apply to murder cases, it can also be used in neglect or abuse cases of children and the elderly. For these cases, the presence of head lice and maggots is the most apparent evidence of unhygienic conditions and untreated injuries on a living person. Moreover, the neglected time frame can also be estimated based on their amounts and development (Volckaert, 2020).

The application of insects in forensics is further extended with the advance in molecular tools. Not only could the species of insects that surround corpses be identified

more precisely, but the host of the insects that have been fed on could also be identified. With the advance in molecular tools, the analysis of human DNA extracted from the carrion community could be conducted. This analysis may become necessary when the source of the insects' food is eager to be known in conditions such as when only maggots, but no cadaver is present at the possible murder scene or when there is an alternative food source other than the cadaver (Amendt et al., 2004).

2.1.1 Insects as storage of hosts' DNA

The first report on using human DNA from maggots in a forensic case was published in 2011 (Li et al., 2011). In this case, an unidentified skull was found, and a decapitated body that had already severely decomposed was found 500 m away. In order to link them together, DNA profiles are needed as evidence. Therefore, the blowfly (*Aldrichina grahami*) maggots were sampled from both the skull and the body for human DNA extraction. As a result, complete DNA profiles could be generated from the extracted DNA samples, and this case showed that human DNA could be recovered from insects gathered from badly putrefy cadavers.

Other than the carrion community, insects such as mosquitoes, sandflies, tsetse flies, blowflies, blackflies, bed bugs and dung beetles consume biological materials, including blood, semen, and faeces; they can also act as storage of human and mammals DNA. The presence of all these insects could be a silver lining in forensic cases where the body, as well as the maggots, could not be found; the found body has been severely traumatized to the extent that generating a DNA profile is not possible or if the primary crime scene has already been cleaned following the disposed of the corpse (Kulstein et al., 2015; Rivers & Geiman, 2017; Durdle, 2020). The DNA preserved in these insects may be sufficiently high quality in profiles generating and providing leads in guiding an

investigation, such as determining if a crime has occurred, linking people to locations and other individuals, and identifying potential perpetrators or victims (Durdle, 2020). For example, DNA profiles have been successfully generated from insects, including mosquitoes, *Aedes aegypti* (Ibrahim et al., 2015) and human lice, *Pediculus humanus capitis* (Pilli et al., 2016).

2.2 Bed bug

Bed bugs are flat, wingless ovals (4–7 mm) insects that belong to a family called Cimicidae (Akhoundi et al., 2020). Their heads are pyramid-shaped, and they have prominent compound eyes, slender antennae, and a long proboscis tucked underneath the head and thorax. The adults are reddish-brown, whereas the immature probably be light yellow (Goddard & Deshazo, 2009).

2.2.1 Bed bug taxonomy

Kingdom: Animalia

Phylum: Arthropoda

Class: Insecta

Order: Hemiptera

Family: Cimicidae

Genus: *Cimex*

Species: *Cimex hemipterus*

Cimex hemipterus, commonly known as the tropical bed bug (Plate 2.1), belongs to the phylum Arthropoda, which encompasses a vast group of invertebrate animals with jointed legs and segmented bodies. Within the arthropods, *C. hemipterus* falls under the class Insecta, which includes insects characterized by their three main body segments (head, thorax, and abdomen), six legs, and usually one or two pairs of wings.

Cimex hemipterus is classified under the order Hemiptera, commonly referred to as "true bugs" or "hemipterans." Hemipterans are characterized by their specialized piercing-sucking mouthparts, which they use to feed on plant or animal fluids. This order includes insects such as aphids, cicadas, leafhoppers, and bed bugs.

The family Cimicidae comprise the genus name of *Cimex*, encompassing several bed bug species. The specific epithet *hemipterus* is derived from the Greek words "hemi" meaning "half", and "pteron", meaning "wing," possibly alluding to the bed bug's partially reduced wings or its belonging to the order Hemiptera (Reinhardt & Siva-Jothy, 2007).

Plate 2.1: Female and male *C. hemipterus* (Cho et al., 2023).

2.2.2 Bed bug biology

Bed bugs are hematophagous insects that feed exclusively on vertebrate blood for survival, growth, and reproduction (Akhoundi et al., 2020). Many cimicids parasitize vespertilionid and molossid bats or swifts and swallows, while *Cimex lectularius* and *C. hemipterus* are closely associated with humans (Reinhardt & Siva-Jothy, 2007). Bed bugs usually stay in the creak and crevices of the bugs-infested house during the daytime and leave their refugia to feed when their host or humans has minimal activity, such as at nighttime when their hosts are sleeping. Adult bed bugs require 10 to 20 min to be fully engorged and feed typically once per week (Reinhardt & Siva-Jothy, 2007).

Bed bugs have a unique method of mating called "traumatic insemination" because the male will insert his intromittent organ or paramere into the female body wall into a specialized organ or spermalege instead of her genitalia. The male sperm is then released into the female's body cavity and migrate to her ovaries to fertilize her eggs (Reinhardt & Siva-Jothy, 2007; Delaunay et al., 2011; Miller et al., 2019).

Female bed bugs lay several eggs, which hatch into first instar nymphs after 4 to 12 days. Bed bugs grow through five nymphal instars before appearing as adults, and each instar stage needs a blood meal to develop to the next (Plate 2.2). The whole lifecycle for a bed bug to develop from an egg into a reproductive adult is approximately 37 days (Benoit, 2011; Lai et al., 2016; Miller et al., 2019).

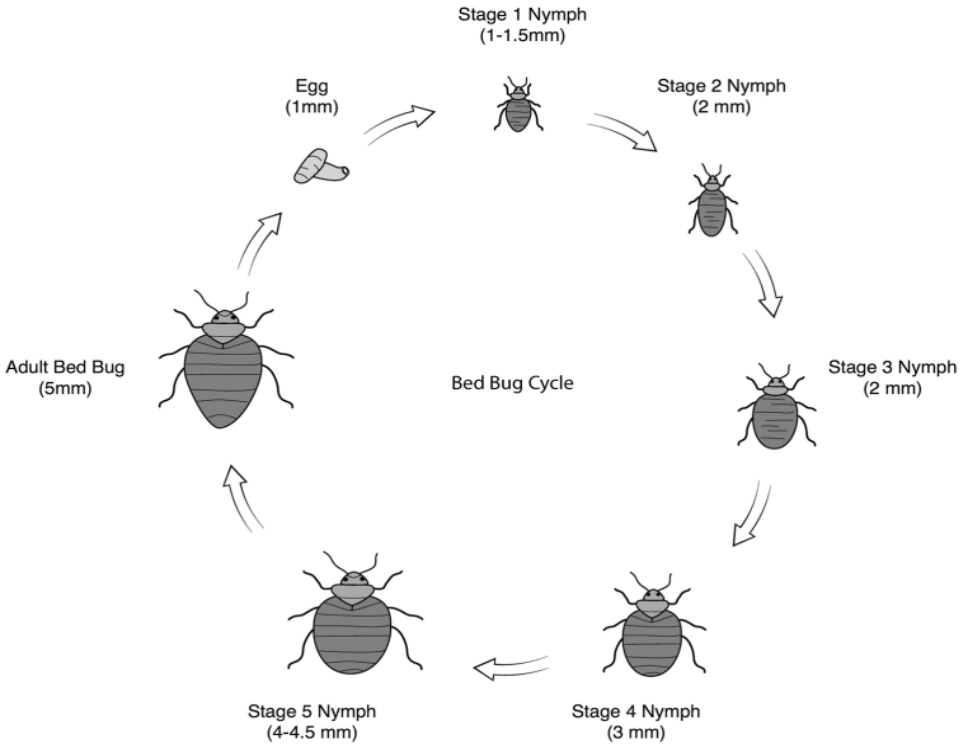

Plate 2.2: Lifecycle of bed bug (Lai et al., 2016).

Insects could disperse either actively or passively. If it distributes actively, the insect moves through its ability (walking or flying). On the other hand, if an insect depends on wind or other organisms for dispersal, it is dispersed passively. For bed bugs, active dispersal was much more limited due to the lack of wings, and the bugs dispersed most of the time passively by hitchhiking on clothing or luggage (Davies et al., 2012; Evison et al., 2018).

2.2.3 Medical importance of bed bugs

Feeding on human blood makes bed bugs regarded as pests, and there are two species of bed bugs that usually feed on humans; *Cimex lectularius*, also known as common bed bugs and *Cimex hemipterus*, also referred to as tropical bed bug (Zorrilla-Vaca et al., 2015; Akhoundi et al., 2020).

Other than blood feeding, bed bugs have also been reported to carry many pathogens, including *Trypanosoma cruzi*, *Rickettsia parkeri*, hepatitis B virus, hepatitis C virus, and methicillin-resistant *Staphylococcus aureus* (Daiter, 1960; Silverman et al., 2001; Villagrán et al., 2008; Goddard & Deshazo, 2009; Lai et al., 2016). However, there are no reports of bed bugs acting as infectious disease vectors in humans, which suggests bed bugs may solely be a phoretic vector for these pathogens (Silverman et al., 2001; Lai et al., 2016).

Although bed bugs cannot transmit disease, they can affect humans in several ways. Bed bug bites could have caused either no response or elicited erythemal reactions and subsequent allergic responses, and the strength of the reaction could be severe to the extent of death. On rare occasions, the bed bugs causing feeding wounds may have allowed other pathogens to enter, rendering secondary infections (Reinhardt & Siva-Jothy, 2007). Other than having adverse effects on health, bed bug infestations can also affect the quality of life, as the consciousness of being bitten could create emotional distress, anxiety, paranoia, and sleeplessness in affected people (Davies et al., 2012).

2.2.4 Economic importance of bed bugs

As the name mentioned, tropical bed bugs, *C. hemipterus*, are reported primarily in the tropics and subtropical regions, such as Africa and Southeast Asia, while common bed bugs, *C. lectularius*, are the most reported species from many countries (especially countries in temperate regions), thus the name "common". However, in recent years, the spread of bed bugs has become global, as species once considered native to specific geographic locations have been found in other parts of the world (Lai et al., 2016). For example, in 2014, *C. lectularius* was detected for the first time in the Chilean province of Magallanes (Faúndez & Carvajal, 2014), while *C. hemipterus* has been found in the

United Kingdom (Davies et al., 2012). The cross-region spread of bed bug populations is probably due to increased global travel and trade, in which people passively carry the bed bugs without recognition (Davies et al., 2012; Lai et al., 2016).

The increase in bed bugs' global population causes distress and economic loss to many sectors, including the tourist, hospitality and poultry industries, schools, hospitals, as well as private and communal households (Reinhardt & Siva-Jothy, 2007; Heukelbach & Hengge, 2009). Among these sectors, the hospitality industry is particularly vulnerable to the infestations of bed bugs because of the high turnover of visitors and luggage. On the same grounds, the initial infestation is also challenging to determine (Penn & Hu, 2020).

The economic loss is due to the cost associated with treating bed bug infestations, such as pest control, including pest treating services, advice and support from professionals, medical expenses for treating staff with allergic reactions, expenses on control equipment and replacing infested infrastructure (Scarpino & Althouse, 2019). As for hotels, infestation causing room closure is a significant loss in terms of revenue (Doggett & Russell, 2008). Moreover, the infestation could damage hotels' social reputation, making potential guests avoid hotels that currently have or have ever had bed bugs (Penn & Hu, 2020). On the other hand, the consequences of infestation in poultry farms include a decrease or even loss in productivity due to afflicted workers and poultry, diminished egg value due to faecal spots from bed bugs, lower egg production and increased feed consumption from affected poultry (Reinhardt & Siva-Jothy, 2007).

2.2.5 Molecular study of bed bugs

Molecular biology is the study performed on the molecular level that chiefly concerns understanding the interactions and regulations between DNA, RNA, and protein biosynthesis (Li & Zhao, 2013). Due to the resurgence, bed bugs have become one of the popular subjects being studied. With the advancing of sequencing techniques, many molecular studies, including sequencing the genome of bed bugs, transcriptome, metagenome, RNA interference (RNAi), and population genetics, that deal with different aspects, such as the biology and physiology of bed bugs, have been conducted.

For example, Benoit et al. (2016) and Rosenfeld et al. (2016) sequenced the draft genome of bed bugs (*C. lectularius*), and genes related to traumatic insemination, insecticide resistance, and chemosensory have been identified. Rosenfeld et al. (2016) also provide transcriptome assembly of bed bugs which is essential to be used in further study to learn about the gene expression of bed bugs under different conditions, thus understanding their function. As for metagenome, which is the common method applied to learn the microbial structure of an environment, has been used by Meriweather et al. (2013) to provide a view of the microbial constituents of bed bugs.

Studies that used RNA interference (RNAi), a technology also known as Post-Transcriptional Gene Silencing (PTGS), can specifically target the gene of interest and silence it, have also been conducted (Almeida & Allshire, 2005; Basnet & Kamble, 2019). For example, Zhu et al. (2012) silenced the cytochrome P450 genes in bed bugs and found out that the insect has increased the susceptibility of pyrethroids, thus determining the xenobiotic function of the cytochrome P450 genes. Population genetics is the study of genetic variations among different populations of an organism, and several studies (Balvín et al., 2012; Booth et al., 2012; Seri Masran & Ab Majid, 2017) have been

conducted to understand the genetic divergence and discrepancy among different bed bug populations.

Although many molecular studies have been conducted on bed bugs, they mainly concentrate on *C. lectularius*. In addition, according to Benoit et al. (2016), bed bugs have many unidentified genes waiting to be discovered. The functions of those identified genes are also yet to be confirmed. All these molecular techniques could be used to learn further about the biology and physiology of bed bugs' change under different conditions, such as the molecular process of bed buds' blood digestion or their microbial changes after blood feeding, which are also interesting to know.

2.2.6 The advantages of bed bugs used in forensic

Blood-feeding insects, such as mosquitoes or fleas that feed on human blood, can inadvertently ingest human DNA. If these insects are found at a crime scene, DNA analysis can be conducted on them to identify the presence of human DNA. Matching this DNA to a suspect's profile can provide evidence linking the individual to the crime.

As bed bugs have a long history of intimate associations with humans (Akhoundi et al., 2020), they could be ideal for forensic investigations. Moreover, in recent decades, there have been drastically increased cases of bed bug infestations in human habitats, including private residents as well as public buildings and transportation (Akhoundi et al., 2020), which could contribute to bed bugs being used in forensics.

In all other respects, bed bugs also have advantages over other insects. First, all its developmental stages are fed on blood, which is better than other blood-feeding insects such as mosquitoes that only female feeds on blood. Second, bed bugs live longer than other insects as they (including nymphs and adults) can persist for up to 12 months

without feeding (Benoit, 2011; Delaunay et al., 2011). Third, bed bugs are not flying insects. The inability to fly makes them indoor inhabitants reside within protective harbourages near their host after accessing their hosts for blood-feeding (Benoit, 2016; Rosenfeld et al., 2016). Therefore, there is a high prospect that their blood meal sources or hosts will be found in proximity to the location of the capture of bed bugs by tiding the DNA profiles (Inbar et al., 2016).

2.3 DNA profile

In a criminal case, matching the suspect's DNA profile with the evidence is always essential in incriminating someone to a crime. However, a DNA match between a suspect and the DNA obtained at a crime scene could be explained in two ways; the DNA legitimately came from the suspect, or the DNA is from someone else with the same pattern of bands. Thus, DNA matching does not necessarily demonstrate that the suspect is guilty but a piece of evidence to be examined along with the other facts in the case (Klug et al., 2019).

DNA evidence was first used in a criminal trial in England in 1986 and in a rape case in Florida, United States, in 1987 (Klug et al., 2019). Since then, DNA typing has revolutionised forensic science, with STR profiling becoming the routine of human identification (Butler, 2005; Nims et al., 2010). Short Tandem Repeats profiling is an analytical DNA technique that amplifies variable microsatellites or STR from a genomic DNA template using PCR, later separates the PCR amplicons on a genetic analyser, and then uses software to analyse the resulting data (Nims et al., 2010).

2.3.1 Short Tandem Repeats (STR) profiling

Short Tandem Repeats, also known as microsatellites, are DNA segments (usually four nucleotides per segment or tetranucleotide) that appear repeatedly and tandemly on the same chromosome. Human genomes contain 5–10% of such repetitive sequences, which vary in size and length among individuals in a population (Hunter, 2006; McClintock, 2014; Vitoševic et al., 2019; Wang et al., 2020). Individuals were then differentiated by the number of copies of the repeat sequence within each STR locus typed.

The PCR products from STR amplification are separated and detected using capillary electrophoresis (CE) (Butler, 2005). As these STR follow a general mendelian inheritance pattern in which the offspring inherits one allele from the mother and the other allele from the father, thus the results display at any typed locus either a single peak denoting homozygosity or two peaks denoting heterozygosity (Panneerchelvam & Norazmi, 2003; Raja et al., 2020) (Plate 2.3). The peaks were then measured using an allelic ladder and marked with the number of repeat units. For instance, a sample containing two alleles, one with 11 and the other with 12 repeat units, would have an "11, 12" genotype (Plate 2.3). This shorthand method of designating the alleles present in a sample makes it easier to compare results from multiple samples.

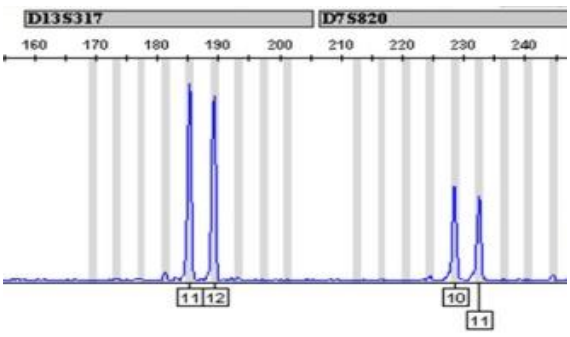

Plate 2.3: STR profile generated using CE with loci D13S317 and D7S820 (Reid et al., 2013).

The DNA profile combines several typed STR markers (commonly used 13 to 16). In DNA typing, the more STR loci being typed, the greater the discrimination value since the likelihood that a single individual has an identical STR profile that possesses the same number of repeat units for all the STR being analysed, with another individual taken at random in the population is extremely rare or not possible (Panneerchelvam & Norazmi, 2003; Hunter, 2006; Klug et al., 2019).

2.3.1(a) STR markers

In November 1997, 13 tetrameric (four base-pair repeat) STRs were developed into a marker panel (called the Combined DNA Index System or CODIS panel) used by the FBI and other law agencies (Butler, 2005; Klug et al., 2019). These 13 STR markers included CSF1PO, FGA, TH01, TPOX, VWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, and D21S11 (Butler, 2005). Establishing a core set of STR loci ensures that all forensic laboratories in the USA and over 25 other countries worldwide can set up uniform DNA databases (Bloom, 1994; Budowle et al., 2001). These loci are distributed over multiple chromosomes. The population frequency of each of these STR alleles has also been measured in many people groups worldwide. The average random match probability of these STR alleles is rarer than one in a trillion among unrelated individuals, thereby providing a high power of discrimination (Butler, 2005; Fang et al., 2011; Klug et al., 2019).

The nomenclature for DNA markers is straightforward. If a marker is part of a gene, the gene name is used in the designation. For example, the STR marker TH01 is from the human tyrosine hydroxylase gene located on chromosome 11, thus "TH". The "01" of TH01 is because the repeat region is located within intron 1 of the tyrosine hydroxylase gene.

DNA markers that fall outside gene regions may be designated by their chromosomal position. The STR loci D5S818 and D16S539 are examples of markers not found within gene regions. In these cases, the "D" stands for DNA. The next character refers to chromosome number - "5" for chromosome 5 and "16" for chromosome 16. The "S" refers to the DNA marker being a single copy sequence. The final number indicates the order in which the marker was discovered and categorised for a particular chromosome. Sequential numbers give uniqueness to each identified DNA marker (Butler, 2005).

2.3.1(b) Y-STR markers

The ability to designate whether a sample originated from a male or female contributor is valuable, especially in sexual assault cases. The most popular method for sex typing is the amelogenin typing system since the DNA encoding gender can be amplified in conjunction with STR analysis.

Several genetic markers have also been identified on the Y chromosome that is distinct from markers on the autosomes and is helpful for human male identification. The Y-STR markers are found in the noncoding region on both arms of the Y chromosome. The Y-STR markers produce a haplotype profile when amplified from male DNA. Such a profile simplifies the interpretation of a mixture containing a male and female contributor by eliminating the female contribution from the amplification profile (McClintock, 2014).

2.3.1(c) Common problem faced in STR profiling

Although high mutational rates at these STR loci lead to extensive polymorphism, making them particularly useful in studying population genetics, the same characteristics make them less stable (Vitoševic et al., 2019). Because they are less stable, polymerase slippage is more likely to occur during DNA replication and causes artefact fragments or stutter peaks that showed on CE, which are one repeat shorter or longer than the original allele at a lower copy number. Stutter fragments can complicate the profile analysis, especially the analysis of imbalanced mixed-source samples, as the amount of DNA in mixture samples from the different contributors could be very different. The true alleles from the minor contributors can be in the same range as peaks generated from alleles of the majors or being misinterpreted as stutter peaks and vice versa (De Knijff, 2019; Oldoni et al., 2019).

STR analysis that involves low template samples could also have problems such as stochastic effects that may result in different alleles being observed at a particular locus and affect the interpretation of an STRs profile (Ziętkiewicz et al., 2012). Another problem of CE is that it does not provide information regarding the underlying base pair variation of the DNA sample; thus, DNA fragments with similar-sized but different sequence compositions could not be detected (Ziętkiewicz et al., 2012; De Knijff, 2019). Severely degraded DNA samples containing only very short DNA template molecules (under 150 bp) may also make STR typing (usually 150–400 bp) unsuccessful (Bianchi & Liò, 2007).

Despite all the mentioned problems, genetic analysis of STR loci is still the most convenient and irreplaceable DNA typing for human identification (Schneider et al., 2004). Thus, STR analysis should be tread lightly from tip to toe and only look for other extended analyses when STR typing fails.

2.3.2 Single Nucleotide Polymorphisms (SNP) profiling

Single nucleotide polymorphisms are the simplest form of DNA variation among individuals (Shastry, 2009). It refers to single-base differences in the DNA sequence, which generally have just two alleles, for example, one allele with guanine and one with adenine, and therefore are not highly polymorphic. However, SNP occur every 1/2000 bases along with a DNA strand through base substitutions, insertions, or deletions at single positions, making them an almost unlimited source of diversity in the genome. Thus, SNP could be analysed through an abundance of potential markers, and the power of discrimination could be very high by typing hundreds of them (Budowle & Van Daal, 2008; Ziętkiewicz et al., 2012; Vitoševic et al., 2019).

The SNP for forensic analyses can be divided into four categories: identity-testing, lineage-informative, ancestry-informative, and phenotype-informative (Budowle & Van Daal, 2008). The function of identity-testing SNP is the same as STR profiling used for individualisation. Nevertheless, SNP have limited discrimination in personal identification compared to STR. Around 50–80 SNP must be typed to achieve the same discriminatory power of typing 13 core STR loci, and typing such a vast number of SNP is more challenging than just analysing 13 STR loci with current technology (Budowle & Van Daal, 2008). Also, the cost of SNP typing will be much higher than currently well-established STR loci and the Combined DNA Index System (CODIS) database. The ability of SNP to resolve samples containing more than one contributor or mixture samples is also much more problematic than STR (Nicklas & Buel, 2008; McClintock, 2014). Therefore, these single-base differences are better used as a virtual extension to routine STR-based DNA profiling (Budowle & Van Daal, 2008; Lou et al., 2011; Ziętkiewicz et al., 2012; Vitoševic et al., 2019).

As for lineage-informative SNP, sets of tightly linked SNP could be used to identify relatives with higher probabilities due to these SNP having low mutation rates, making them change less likely over generations. These SNP are crucial concerning inheritance cases or kinship analyses and family reconstructions for missing persons and situations where no direct reference sample may be available (Budowle & Van Daal, 2008; Vitoševic et al., 2019).

Ancestry-informative SNP are human polymorphisms that substantially exhibit allele frequency differences among populations. These SNP can provide information about samples' ancestry, which may help predict a perpetrator's ethnic origin to aid criminal investigations (Budowle & Van Daal, 2008; McClintock, 2014).

The last one is phenotypic-informative SNP, which are used to predict a sample or person's phenotypes. The most apparent characteristics of an individual's appearance are colouring, height, and facial features. All these traits are highly heritable, suggesting that the genetic polymorphisms responsible for these different phenotypic traits could be determined. So far, a set of SNP have been developed known as HIrisPlex-S, and they could be used to predict a person's skin, hair, and eye colours (Budowle & Van Daal, 2008; McClintock, 2014).

Therefore, although SNP is not efficient enough to replace STR, it offers several advantages over STR and could help shed some light in some cases. For example, when there is no suspect in a case, SNP that describes a person's phenotypic traits will enable a genetic prediction of appearance and help identify the perpetrator. In addition, the appearance prediction also helps to confirm or refute the description of suspects from eyewitnesses, which is not always reliable. Unlike STR markers that concentrate on personal information, SNP markers are mostly population-oriented, which can contribute

vital information on biogeographic ancestry that STR cannot deliver (Vitoševic et al., 2019). Moreover, the amplicons of SNP could be reduced to as short as 50-60 bp, which makes these SNP markers paramount when dealing with compromised or degraded DNA samples (Budowle & Van Daal, 2008; Ziętkiewicz et al., 2012; Vitoševic et al., 2019).

2.4 Digestive system of bed bugs

Blood containing human DNA is essential in forensic uses, but its presence within the bed bugs could be limited, probably due to digestion. As the volume of blood obtained could affect the DNA analysis, it is vital to understand bed bugs' blood meal processing. Blood is not a perfect meal to be solely fed on as it mainly contains only proteins, can produce enormous amounts of haem after oxidising, which could be dangerous, and has a shortage of some vitamins. Therefore, as an insect that exclusively feeds on blood, other than their digestive system, they may also require bacterial endosymbionts to survive (Ribeiro & Arca, 2009).

2.4.1 Microbiome, metagenomics and metabarcoding

The term microbiome refers to the entire community of microorganisms, including bacteria, archaea, viruses, phages, fungi, and protozoa, that exist within any ecosystem, such as plants, insects, animals, humans or environmental samples. However, most microbiome studies involve only bacteria and archaea (Handelsman et al., 2007).

Traditionally, to study the microbiome, the species must be isolated from the ecosystem and cultured separately. The introduction of high-throughput DNA sequencing and advanced computing capabilities make the metagenomics and metabarcoding

approaches are now possible (Handelsman et al., 2007). Metagenomics is a cultivation-independent analysis of the genetic information of the collective genomes of the microbes within a given environment based on its sampling (Izard, 2015; Roumpeka et al., 2017), while metabarcoding sequences only the barcode from all genomes presents in a sample instead of sequencing the whole genome. For instance, 16S ribosomal RNA (rRNA) for bacteria and archaea or 18S for fungi.

Metabarcoding is popular as only a few hundred base pairs of a particular sequence that could be found in a genome collection were sequenced for taxonomic profiling. This way not only reduces the size of the reference genomes database but also decreases the computational time and hardware requirements and is more cost-effective (Scholz et al., 2015). Moreover, read counts produced per sequencing of metabarcoding are massive and could be used to quantify the microbes. Another advantage is that uncultured microbes could also be sequenced and identified (Bukin et al., 2019).

2.4.1(a) Bacterial community studies in insects using metabarcoding

It is well-known that microorganisms are harboured by insects, especially bacteria, and live symbiotically with their host (Malacrinò et al., 2018). Symbiotic bacteria often play significant roles in metabolic or nutritional functions in insects. For example, they provide nutrients, help digestion, tolerate extreme temperatures and help their host survive under a limited diet (Sudakaran et al., 2012; Naaz et al., 2020). Moreover, insects with microbial symbionts also exhibit a higher survival rate against insecticides and fungal pathogens (Scarborough et al., 2005; Kikuchi et al., 2012). Thus, many studies have focused on identifying the bacteria species and understanding their role in an insect.

Before metabarcoding, studies of insect bacteria were focused on single endosymbionts and their interactions with the host. However, many insects are inhabited

by a complex bacterial community (Sudakaran et al., 2012). In order to identify and characterize the whole bacterial community, metabarcoding using a 16S rRNA marker is the most common way.

The 16S rRNA gene is part of the small ribosomal subunit (SSU) and has been used for taxonomic studies of prokaryotic species universally since the 1960s and 70s as this gene is present in all microbes (Scholz et al., 2015; Santos et al., 2020). Its length is approximately 1600 base pairs (bp) and includes nine hypervariable regions (V1-V9) that can be used as taxonomic tags (Scholz et al., 2015; Bukin et al., 2019).

Most metabarcoding works, such as on fruit flies and bees (Malacrinò et al., 2018; Voulgari-Kokota et al., 2019), use the Illumina platform, which usually produces paired-end reads up to 150 bp (Bukin et al., 2019). All these 16S rRNA genes or gene fragments have fuelled and made it possible to build databases using this marker, such as the Greengenes database (DeSantis et al., 2006), the Ribosomal Database Project (RDP) (Cole & Tiedje, 2014), and SILVA (Quast et al., 2012), and continue to extend the insect-bacterial community studies (Ogier et al., 2019).

2.4.2 Transcriptome analysis using RNA-Seq

RNA-Sequencing (RNA-Seq) is now widely used in almost every field of biological studies and has extended the view of transcriptomic complexity in different species (Cheng et al., 2018). The transcriptome is the complete set of transcribed sequences in a cell, including messenger RNA (mRNA), ribosomal RNA (rRNA), transfer RNA (tRNA), and noncoding regulatory RNA (He, 2015).

RNA-Seq is usually used to detect functional genes as it only sequences the coding genes. Thus, one of the most common aims of transcriptome analysis is to identify