DEVELOPMENT OF NOVEL MICROSATELLITE MARKERS AND ASSESSMENT OF POPULATION GENETICS IN THE TROPICAL BED BUG, *Cimex hemipterus* F. (HEMIPTERA: CIMICIDAE)

WAN NUR FATANAH BINTI WAN MOHAMMAD

UNIVERSITI SAINS MALAYSIA

2020

DEVELOPMENT OF NOVEL MICROSATELLITE MARKERS AND ASSESSMENT OF POPULATION GENETICS IN THE TROPICAL BED BUG, *Cimex hemipterus* F. (HEMIPTERA: CIMICIDAE)

by

WAN NUR FATANAH BINTI WAN MOHAMMAD

Thesis submitted in fulfilment of the requirements
for the degree of
Master of Science

June 2020

ACKNOWLEDGEMENT

All praise and worship to Allah the Almighty for the strength and His blessings in completing this degree of Masters of Science. I owe my highest gratitude to my supervisor Dr. Veera Singham a/l K. Genasan for his supervision, encouragement and support throughout the completion of this study. It is an honor to be under his dedicated supervision. Not forgotten, my appreciation to my co-supervisor Mdm. Wan Nurainie binti Wan Ismail for her advices and encouragement. A special thanks I would like to convey to Universiti Sains Malaysia (USM) especially Centre for Chemical Biology (CCB) for providing me with all the necessary equipment and facilities. This special thanks is also conveyed to the lab members Fazli, Li Shen, Aliza, Yana, Diana, Syima, Fiza, Kak Ila, Kak Farah, Kak Syak, Kak Komala, Kak Min, Jess, Sam, Boon Tiong and Yeap. Not to forget the administrative department of CCB for being helpful, Mr. Zulkifli Md. Yusoff, Mdm. Nurul Amira Mohd Ali, Mdm. Tengku Zalina Tengku Ahmad and Mdm. Nur Amrina Rosyada Amran. Besides, I would like to express my fond thanks to my beloved mother, father and sisters who always love, care and support me in many ways. They are the one who always being there during my ups and downs. Their priceless love and support are hardly to repay, and all my hard work are truly for them. Last but not least, I would like to extend my sincere appreciation to Najmi Naim for his valuable encouragement and support that made this journey tolerable. This study would never have been accomplished without their pure love and patience.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

AMOVA	Analysis of molecular variance
APS	Ammonium persulphate
A_R	Allelic richness
BAPS	Bayesian Analysis of Population Structure
bp	Base pair
COI	Cytochrome oxidase I
DNA	Deoxyribonucleic acid
EDTA	Ethylenediamine tetra-acetic acid
EM	Expectation-Maximization
EtBr	Ethidium bromide
F_{IS}	Level of inbreeding
F_{ST}	Genetic differentiation
g	Gram
gDNA	Genomic DNA
h	Hour
H_E	Expected heterozygosity
H_O	Observed heterozygosity
min	Minute
ml	Mililitre
mPCR	Multiplex PCR
MPSS	Massively Parallel Signature Sequencing
mtDNA	Mitochondrial DNA
N_A	Allelic richness

NCBI	National center for biotechnology information
NGS	Next generation sequencing
NJ	Neighbour joining
nm	Nanometer
OD	Optical density
P	*P*-value to check significant of the analysis
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PIC	Polymorphic information content
r	Relatedness
rpm	Rotation per minute
s	Second
sdH_2O	Sterile distilled water
TAE	Tris-acetate-EDTA
TBE	Tris-borate-EDTA
TEMED	Tetramethylethylenediamine
µl	Microliter
µM	Micromolar
UV	Ultra violet
V	Volt
w/v	Weight/volume
16S rDNA	16S ribosomal DNA

PEMBANGUNAN PENANDA BARU MIKROSATELIT DAN PENILAIAN GENETIK POPULASI DALAM PEPIJAT TROPIKA, *Cimex hemipterus* F. (HEMIPTERA: CIMICIDAE)

ABSTRAK

Infestasi pepijat semakin meningkat di Malaysia. *Cimex hemipterus* (Fabricius) ialah ektoparasit penghisap darah pada waktu malam yang dominan di rantau tropika. Walaupun kadar infestasi pepijat tropika adalah membimbangkan, namun kajian mengenai populasi genetik spesis ini adalah kurang. Bagi menyimpulkan permasalahan ini, matlamat kajian kami ialah untuk membangunkan dan mencirikan penanda baru mikrosatelit daripada genom *C. hemipterus* di samping menilai populasi genetik dalam Malaysia dan Singapura berdasarkan penanda mikrosatelit yang baru dibangunkan daripada kajian ini dan gen asid deoksiribonukleik mitokondria (mtDNA) yang dinamakan sitokrom oksidase I (COI) dan 16S asid deoksiribonukleik ribosom (16S rDNA). Penanda baru mikrosatelit telah dibangunkan menggunakan teknik jujukan generasi (Roche 454 pyrosequencing) dalam kajian ini. Sebanyak 52,119 bacaan telah didapati dengan 1,821 (3.49%) bacaan mengandungi mikrosatelit jenis pengulangan motif di-, tri- dan tetranukleotida. Lapan penanda mikrosatelit telah digunakan untuk menilai populasi genetik *C. hemipterus* di Malaysia dan Singapura. 351 individu *C. hemipterus* untuk seluruh lapan lokus telah digenotipikkan. Kesemua populasi menunjukkan heterozigositi yang defisit, kecuali BMV (heterozigositi diperhatikan, H_O = 0.71; heterozigositi dijangkakan, H_E = 0.62). Walau bagaimanapun, tahap kepelbagaian genetik yang rendah hingga sederhana diperhatikan di semua populasi (kekayaan alel, A_R: 1.85 - 3.75; bilangan alel pada setiap lokus, N_A = 1.88 - 3.88).

Pengagihan kekerapan alel menunjukkan bahawa populasi pepijat katil tropika dalam kajian semasa mengalami kesesakan genetik baru-baru dan turut disokong oleh nilai indeks Garza-Williamson (GW). Tahap pembiakbakaan yang sederhana didapati dalam 18 populasi *C. hemipterus* di Malaysia dan Singapura (F_{IS} = -0.15 - 0.46), kecuali dalam BMV (F_{IS} = -0.15). Ini berkemungkinan berlaku disebabkan tahap pengenalan genetik berganda daripada sumber tidak berkaitan dalam populasi atau melalui penyebaran aktif pada jarak pendek seperti antara pangsapuri bersebelahan dalam bangunan yang sama. Analisis keterkaitan genetik (r = 0.124 - 0.844) menunjukkan bahawa peristiwa pengawanan dalam kalangan kerabat yang tinggi adalah perkara biasa dalam sesetengah populasi *C. hemipterus*. Di samping itu, 17 kelompok telah dikelompokkan dalam populasi spesis ini menggunakan kluster 'Bayesian'. Pemerhatian ini mungkin berlaku kerana keterkaitan genetik yang tinggi dan berlakunya kesesakan genetik dalam setiap populasi. Analisis Struktur Populasi secara Bayesian (BAPS) turut memperlihatkan kehadiran admikstur genetik yang signifikan (P < 0.05), yang menggambarkan bahawa penyebaran pepijat katil tropika boleh berlaku melalui penyebaran pasif di antara lokasi geografi yang dekat atau boleh berlaku disebabkan penduduk di antara dua atau lebih populasi berpotensi berkongsi sumber yang sama melalui aktiviti yang dikendalikan oleh manusia. Hasil daripada analisis perubahan molekular genetik (AMOVA) menggambarkan bahawa 28.50% (P <0.001) variasi genetik dikaitkan dengan kepelbagaian populasi, 18.67% (P < 0.001) kepada individu dalam populasi dan 52.83% (P <0.001) kepada individu. Walaupun perbezaan signifikan antara populasi adalah jelas, yang mana menunjukkan tahap struktur yang tertentu dalam kalangan populasi *C. hemipterus*, variasi dalam individu populasi juga sangat tinggi. Hal ini menunjukkan substruktur genetik berpotensi berlaku dalam kajian semasa. Di samping itu, penilaian genetik

berdasarkan gen mtDNA menunjukkan ketiadaan variasi genetik yang jelas antara semua populasi dengan ketiadaannya heteroplasmi tidak seperti sebelumnya diperhatikan dalam kalangan populasi *Cimex lectularius*. Dapatan ini menunjukkan bahawa kawasan mtDNA ini mungkin kurang bermaklumat dalam menjelaskan populasi genetik *C. hemipterus*. Kajian masa depan yang mensasarkan penanda molekul lain akan berguna dalam memahami lagi struktur genetik populasi kriptik *C. hemipterus*.

DEVELOPMENT OF NOVEL MICROSATELLITE MARKERS AND ASSESSMENT OF POPULATION GENETICS IN THE TROPICAL BED BUG, *Cimex hemipterus* F. (HEMIPTERA: CIMICIDAE)

ABSTRACT

Bed bug infestations are on the rise in Malaysia. The tropical bed bug, *Cimex hemipterus* (Fabricius) is a nocturnal blood-sucker that is predominant in the tropical region. Despite its alarming infestation rate, the information on its population genetics remains scarce. To infer this, the study is aimed at developing and characterizing novel microsatellite markers from the genome of *C. hemipterus* and to assess population genetics within Malaysia and Singapore based on the newly developed microsatellite markers and mitochondrial DNA (mtDNA) genes namely cytochrome oxidase I (COI) and 16S ribosomal deoxyribonucleic acid (16S rDNA). Here, the next generation 454 Roche pyrosequencing technique was used to develop the novel microsatellite markers. A total of 52,119 reads with 1,821 (3.49%) reads containing microsatellites with di-, tri- and tetra-nucleotide repeat motifs were obtained. Eight primers were used to assess the population genetics of *C. hemipterus* within Malaysia and Singapore. 351 individuals of *C. hemipterus* were genotyped across eight polymorphic microsatellite loci. All populations in this study showed heterozygosity deficit, except BMV (observed heterozygosity, H_O = 0.71; expected heterozygosity, H_E = 0.62). Nonetheless, a low to moderate level of genetic diversity was observed across all populations (allelic richness, A_R : 1.85 - 3.75; the number of alleles per locus, N_A = 1.88 – 3.88). The allele frequency distribution displayed that the populations of the tropical bed bug in this study were experiencing recent bottleneck and it was supported by the value of Garza-Williamson (GW) index. A

considerable level of inbreeding was found within 18 populations of *C. hemipterus* within Malaysia and Singapore (F_{IS} = -0.15 – 0.46) with an exceptional in BMV (F_{IS} = -0.15). This is maybe due to the high level of unrelated multiple introductions in the population or through active dispersal at short distances such as between adjacent apartments in the same building. The genetic relatedness assessment (r = 0.124 - 0.844) indicated that consanguineous mating events are common in some populations of *C. hemipterus*. In addition, 17 clusters were clustered in this study using Bayesian clustering which may happen due to the high relatedness and occurrence of inbreeding within each population. Bayesian Analysis of Population Structure (BAPS) also displayed the presence of significant admixture (P < 0.05), which depicted that dispersal can occur through passive dispersal between geographically proximate locations or where the residents between two or more populations potentially share a common source of infesting bed bug source population which is possible via human-mediated activities. The analysis of molecular variances (AMOVA) depicted that 28.50% (P < 0.001) of genetic variation was attributable to among population diversity, 18.67% (P < 0.001) to among individuals within populations and 52.83% (P < 0.001) to within individuals. Despite the significant differentiation between populations which suggesting a certain degree of structure among the populations of *C. hemipterus*, the significantly high variation within individual was also suggesting potential substructure within the population. Besides, the genetic assessment using mtDNA genes resulted in no obvious genetic variation between all populations with the absence of heteroplasmy, unlike previously observed in the sibling species *Cimex lectularius*. This suggested that these regions of mtDNA are probably less informative for population genetics of *C. hemipterus*.

Future studies targeting different regions would be useful in further understanding the cryptic population genetics structure of *C. hemipterus*.

CHAPTER 1

INTRODUCTION

The eradications of bed bug are expensive due to current challenges in control strategies followed by lawsuits among the operators of the hospitality industry and individuals from private residence. The bed bug is a nocturnal blood-sucking ectoparasite that has now emerged as important public health pest globally. It was reported by Usinger (1966) and Panagiotakopulu and Buckland (1999) that this insect has been causing physical, psychological and medical complications in humans (Booth et al., 2012; Delaunay et al., 2011; Goddard & deShazo, 2009) for at least 3350 years. Generally, two species of bed bug that are closely associated to human are *C. lectularius* (L.) and *C. hemipterus* (F.). A significant morphologically difference can be observed at their prothorax. *C. hemipterus* has squared off and deeper prothorax near the eye whereas *C. lectularius* has rounded prothorax portions near the eye, more enlarged laterally and the edge are more flattened (Ghauri, 1973). As stated by Service (1996), *C. hemipterus* is normally encountered in tropical and subtropical countries and its growing number of infestations in Southeast Asia has gained public attention since a decade ago (How & Lee, 2010; Khan & Rahman, 2012;Tawatsin et al., 2011). Nonetheless, *C. hemipterus* can also be found in areas with much cooler climate due to their ability to be passively disperse via travellers carrying infested items.

Previously, bed bug infestations has been correlated with economically poorer regions (Kells, 2006), however, now it can be encountered in public accommodations (Harlan, 2006; How & Lee, 2010; Khan & Rahman, 2012; King et al., 1989; Potter, 2006) such as hotel, movie theater, airport, family units (Boase, 2008; Criado et al.,

2011; Romero et al., 2007), apartments (Booth et al., 2012), rooming houses, offices, health care facilities and college dormitories (Hwang et al., 2005; Williams & Willis, 2012). The attention towards this species was generally disappeared at the end of World War II since the effectiveness of modern insecticides and hygiene in industrialized and developed countries have improved (Mumcuoglu, 2010). Nevertheless, increased in global movement of human (Kells, 2006; Potter, 2006; Williams & Willis, 2012), evolution of insecticide resistance (Doggett et al., 2004; Potter, 2006; Romero et al., 2007) and changes in patterns of insecticide usage have also led to the resurgence of this pest species. The evolvement of the insecticide usage has changed from premise-wide use of broad-spectrum insecticides to a more selective control tactics for other urban pests (Doggett et al., 2004; Potter, 2006). Thus, the resurgence of this pest has heightened the sensitivity of the general public towards the risk potential of bed bugs infesting homes and other residential units as well as public facilities (Boase, 2008; How & Lee, 2010; Williams & Willis, 2012).

Bed bug is an insect that blood-fed on human, domestic animals, birds and bats (Goddard & deShazo, 2003). Even though this pest is not established as a vector of diseases to human (Blow et al., 2001; Goddard & deShazo, 2003; Pinto et al., 2007; Usinger, 1966), few recent studies identified that bed bug has the potential to transmit a number of human pathogens in laboratory settings using mice (Delaunay et al., 2011; Webb et al., 1989). The bite of the bed bug can cause medical nuisance to human, which includes macules, papules, vesicles, wheals, bullae or nodules (Criado et al., 2011; Doggett & Russell, 2007; Goddard & deShazo, 2003). The systemic reactions such as urticaria, asthma and anaphylaxis (Criado et al., 2011; Goddard & deShazo, 2003) can also be triggered in more sensitive individuals. Bed bug infestation also affects the emotional health and well-being of individuals because of

anxiety, sleep loss (Goddard & deShazo, 2003; Pinto et al., 2007), psychological trauma (Doggett & Russell, 2007; Hwang et al., 2005; Usinger, 1966) and social ostracism (Ryckman, 1985; Ryckman et al., 1981; Thomas et al., 2004).

To date, lack of attention is given to the population genetics and breeding structure of bed bug (Reinhardt & Siva-Jothy; 2007), in which these aspects are important to understand the infestation dynamics of this insect. So far, the information on genetic diversity, phylogenetics, population dynamics and dispersal is well-documented for *C. lectularius* but not in *C. hemipterus* (Akhoundi et al., 2015; Booth et al., 2012; Narain et al., 2015; Saenz et al., 2012; Szalanski et al., 2008). On top of that, the de novo microsatellite markers which were used to elucidate the infestation dynamics in *C. lectularius* were found to be highly species-specific and attempts at cross-amplifying these markers in *C. hemipterus* was not successful. Therefore, it is important to develop and characterize microsatellite markers from the genome of *C. hemipterus* as it can provide a new insight in elucidating the population dynamics of this pest species at molecular level. The use of mtDNA in population genetics able to track the maternal history (Besansky et al., 2003) and biogeographical history of the target organism (Avise et al., 2000; Soltis et al., 2006). It can also be used in detecting heteroplasmy within a cell or individual (Robinson et al., 2015). Nevertheless, the mtDNA genes show drawbacks in this study which may be due to the low levels of genetic variability. This is different to microsatellite markers that sensitive enough for population genetics assessment. To this extent, this study documents the population genetics and breeding structure of *C. hemipterus* through the development of novel polymorphic microsatellite markers and mtDNA analyses of bed bug populations in in Malaysia and Singapore.

Therefore, the objectives of this study are:

1. To develop and characterize novel microsatellite markers from the genome of *C. hemipterus* through next generation sequencing.
2. To assess genetic diversity, breeding structure, genetic differentiation and genetic structure of *C. hemipterus* within Malaysia and Singapore as inferred from the developed polymorphic microsatellite markers.
3. To assess phylogenetics of *C. hemipterus* within Malaysia and Singapore and Japan as inferred from mitochondrial gene, COI and 16S rDNA.

CHAPTER 2

LITERATURE REVIEW

2.1 The Bed Bug

2.1.1 The Resurgence of Bed Bug

Bed bug infestations occur throughout the world. The consequences of bed bug infestation are economically damaging since the infestation can cause millions of dollars in losses through both treatments and lawsuits within the hospitality industry and other residential settings (Dogget & Russel, 2007; Reinhardt & Siva-Jothy, 2007). Its resurgence has triggered an increase in awareness of this pest (Boase, 2008; How & Lee, 2010; Reinhardt & Siva-Jothy, 2007; Romero et al., 2007; Williams & Willis, 2012).

A finding by Panagiotakopulu and Bucklands (1999) mentioned that bed bugs have been long related with human since the discovery of bed bug samples found in Egyptian tombs dating back about 3000 years ago. Formerly, the infestations of this pest have been correlated with economically poorer regions (Kells, 2006), nonetheless, now it can be encountered basically anywhere (Booth et al., 2012; Harlan, 2006; How & Lee, 2010; Hwang et al., 2005; Khan & Rahman, 2012; King et al., 1989; Potter, 2005; Williams & Willis, 2012). Several reasons were hypothesized due to this resurgence which include increased in global movement of human (Kells, 2006; Potter, 2005; Williams & Willis, 2012), evolution of insecticide resistance (Doggett et al., 2004; Potter, 2005; Romero et al., 2007) and changes in patterns of insecticide usage from premise-wide use of broad-spectrum insecticides

to more selective control tactics for other urban pests (Doggett et al,. 2004; Potter, 2005). Furthermore, the increase in the exchange of second-hand furniture was also considered influenced the possibilities of the resurgence.

The bed bug is suggested to emerge from the related insect families that blood fed on bats through evolutionary history (Balvin, 2012; Johnson, 1941) and it is classified under order Hemiptera, suborder Heteroptera and family of Cimicidae (Goddard & deShazo 2003; Delaunay et al., 2011). The family of Cimicidae is consisting of nearly 100 species parasitizing mostly bats and birds. Three species of bed bug that are closely related to human and bats (Ghauri, 1973) are C. *lectularius*, *Leptocimex boueti* (Brumpt, 1910) and *C. hemipterus*. Nevertheless, *C. lectularius* and *C. hemipterus* are commonly encountered. *C. lectularius* can be found in temperate climates including parts of the United States whereas *C. hemipterus* is found mainly in tropical and subtropical regions (Mallis, 1990). Even so, both can be found in every type of climate due to their passive movement (Dogget et al., 2004; Williams & Willis, 2012).

Koganemaru et al. (2017) reported that the common bed bug has showed the pyrehthroid resistance which is due to the penetration resistance. This is happening due to the thicken and or remodel of the bed bug cuticle which resulted in reducing the insecticide penetration rate (Dang et al., 2017). This could prevent or slow the insecticides from rwching the target sites on nerve cells. On top of that, the tropical bed bug has also developed a high level of pyrethroid and carbamate resistance (Punchihewa et al., 2019).

There are distinct differences in morphology of *C. lectularius* and *C. hemipterus*. In comparison to *C. lectularius*, *C. hemipterus* has squared off and deeper prothorax near the eye whereas (Figure 2.1) *C. lectularius* has rounded prothorax portions near the eye (Figure 2.1), more expanded laterally and the extreme margins are more flattened (Ghauri, 1973).

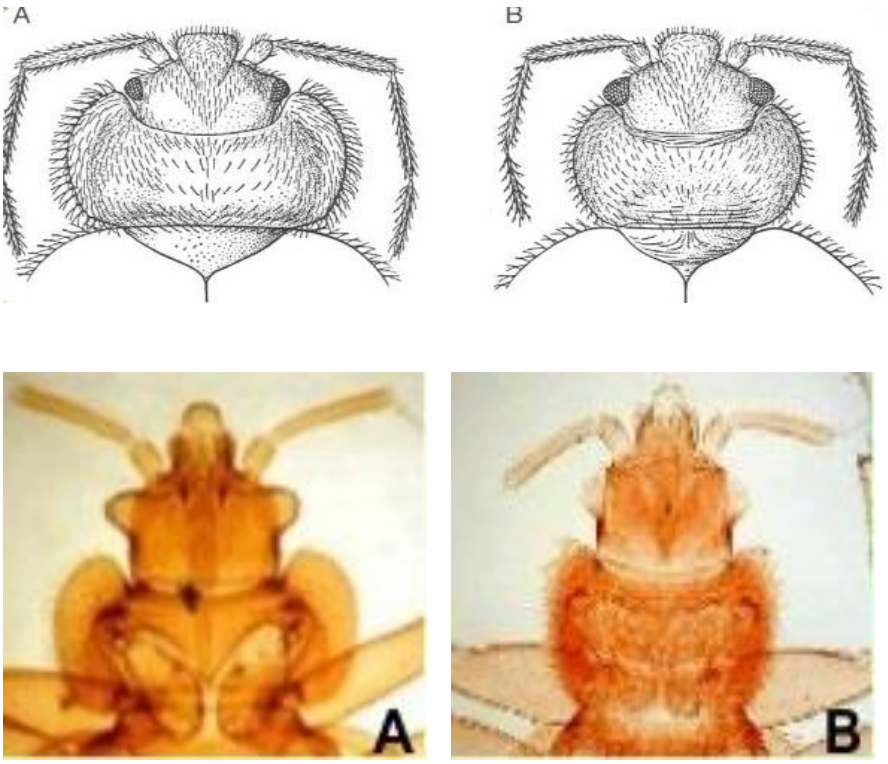

Figure 2.1 Head and prothorax of adult bed bugs. (A) *Cimex lectularius*; (B) *Cimex hemipterus* (Smart, 1943).

2.1.2 Morphology, Behavior and Ecology of Bed Bug

Bed bugs are sometimes mistaken for ticks, cockroaches, carpet beetles or other household insects. This pest feeds solely on the blood of human and warm-blooded animals. Generally, the body of the bed bug is small, oval-shaped, flattened dorso-ventrally when unfed, pale yellow or brown in color, but became reddish brown after a blood meal (Reza & Rahman, 2012).

Sensory towards warmth and carbon dioxide production in bed bugs are efficient. This enables them to easily migrate towards humans for food supply (Williams & Willis, 2012). Its body that previously brownish and flat before blood fed will be elongated and appears more reddish once they are blood fed (Williams & Willis, 2012). An engorged bed bug capable to increase its size in length by 30% to 50% of its original body. Apart from that, the weight can also change by 150% to 200% after a blood meal (Goddard & deShazo, 2003). The bed bug has two separate elongated stylets. Each stylet has a different role during blood feeding. One stylet responsible in withdrawing the blood without clotting (Wilson, 2011), whereas the other stylet injects the saliva that results in their feeding pain-free to the host (Huntley, 1999; Williams & Willis, 2012; Wilson, 2011). Bed bugs can live up to 12 months in a colder environment although in the absence of blood source from hosts (Delaunay et al., 2011; Williams & Willis, 2012). Cimicids are unique parasites among obligate blood feeders because they do not remain on the host to complete their life cycle. After feeding, they move to local areas and hide in places where their host is resting, such as a bed or a couch (Williams & Willis, 2012).

Bed bugs do not fly and jump as fleas do, but they can crawl rapidly over floors, walls, ceilings and other surfaces. A bed bug can disperse both actively and passively, however, the active dispersal is limited. It is capable to crawl up to 20 feet from their harborage to sleeping hosts, yet the movement is very limited (Johnson, 1941; Mallis, 1990). Passive dispersal through human-mediated movement is the common mode of dispersal for bed bug (Loudon 2017). This is because bed bugs are efficient hitchhikers and can be easily transported into dwellings on luggage, clothing, beds, furniture and other belongings. This is a particular risk for hotels and apartments, where the turnover of occupants is constant. Both *C. lectularius* and *C. hemipterus* showed different climbing abilities. *C. hemipterus* has a better vertical friction forces which leads to the better success escape rates compared to *C. lectularius* (Kim et al., 2017). It was suspected that the differences in climbing ability between the two species must be related to variations in the tibial pad.

This pest does not inhabit the visible surfaces of beds, couches or chairs. Rather, they hide in crevices and cracks, or near the seams of chairs, mattresses, in wallpaper, between couch cushions and into the folds of curtains (Williams & Willis, 2012). How and Lee (2010) reported that this species abundant in three most common locations of infestations which are the bedding (31.1%), followed by the headboard (30.3%) and cracks and crevices surrounding the baseboard, wall or floor (23.5%). Infestations can be identified by small specks of bedbug remnants, including their feces and blood from their hosts and cast skins left behind on bed linens when the bedbugs molt (Cleary & Buchana, 2004).

2.1.3 Biology and Reproduction of Bed Bug

Bed bug reproduces through traumatic insemination. Traumatic insemination is a sexual process where the male pierces the female's abdominal wall with his external genitalia and inseminates into her body cavity (Usinger, 1966). According to Reindhart and Siva-Jothy (2007), it is revealed that each female undergoes approximately five traumatic inseminations after each feeding. This is because the male bed bugs tend to direct their sexual activity towards females that have recently fed. The sexual activity involves sperm deposited via male's intromittent bypassing the genital tract and piercing the cuticle (Delaunay et al., 2011). Then, the sperm is injected into the female organ that directs the migration of sperm called mesospermagele. The probability of survival in female bed bugs during copulation are low. In spite of that, each fertilized adult female produces 200 to 500 eggs in her lifetime, on average (Mallis, 1990).

The eggs hatch in four to ten days on average, leading to emerging of immature bed bug with 1 to 3 mm long. The immature bed bugs or nymphs resemble the adults but are smaller and lighter in color. The development of nymph consists of five instars. Blood meal is necessary for the nymphs to develop into the next instars. Each instar will take three to seven days before reaching the next stage until adulthood if the environment is favorable. This quick process is a piece of evidence on how a previously non-infested home may undergo an exponential multiplication of bed bugs within a month (Williams & Willis, 2012). Nymphs of a bed bug can be as small as 1.5 mm and its body color are yellow to translucent before feeding. But, the body of the nymphs turns bright red immediately after feeding (Williams & Willis, 2012).

2.1.4 Bed Bug as A Nuisance in Human Daily Life

The bed bug is a blood-sucking insect that infested on humans, domestic animals, birds and bats (Goddard & deShazo, 2003), but preferentially human. Although there is no evidence of bed bugs classified as a vector species in human (Blow et al., 2001; Goddard & deShazo, 2003; Pinto et al., 2007; Usinger, 1966), few studies identified that it has a high potential to transmit a number of human pathogens (Delaunay et al., 2011; Webb et al., 1989). This insect which its mouthparts are modified to pierce the skin of the host and suck blood with an elongated beak (Khan & Rahman, 2012) can cause a medical nuisance to human. A typical "bite" by bed bug of roughly three to 10 minutes may trigger intense dermatological reactions (Figure 2.2, Figure 2.3) which are including macules, papules, vesicles, wheals, bullae or nodules (Criado et al., 2011; Doggett & Russell, 2004; Goddard & deShazo, 2003). An additional systemic reaction like urticaria, asthma and anaphylaxis may also occur (Criado et al., 2011; Goddard & deShazo, 2003). This incident may also lead to a significant impact on the emotional health and well-being of individuals because of anxiety, sleep loss (Goddard & deShazo, 2003; Pinto et al., 2007), psychological trauma (Doggett & Russell, 2004; Hwang et al., 2005; Usinger, 1966) and social ostracism (Ryckman, 1985; Ryckman et al., 1981).

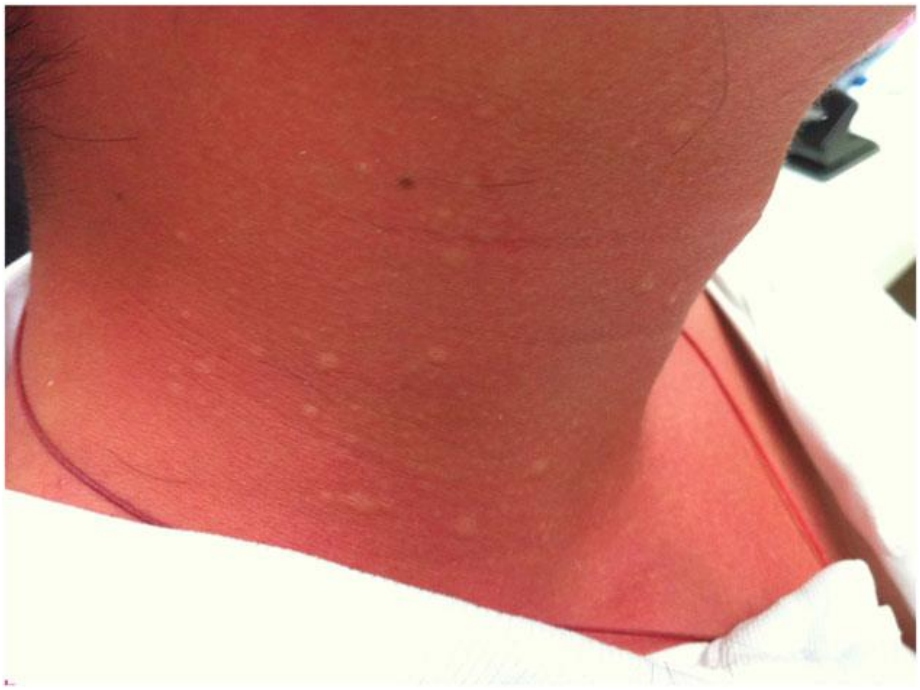

Figure 2.2 Hypopigmented macules over the neck region cause by the bed bugs (Minocha et al., 2016).

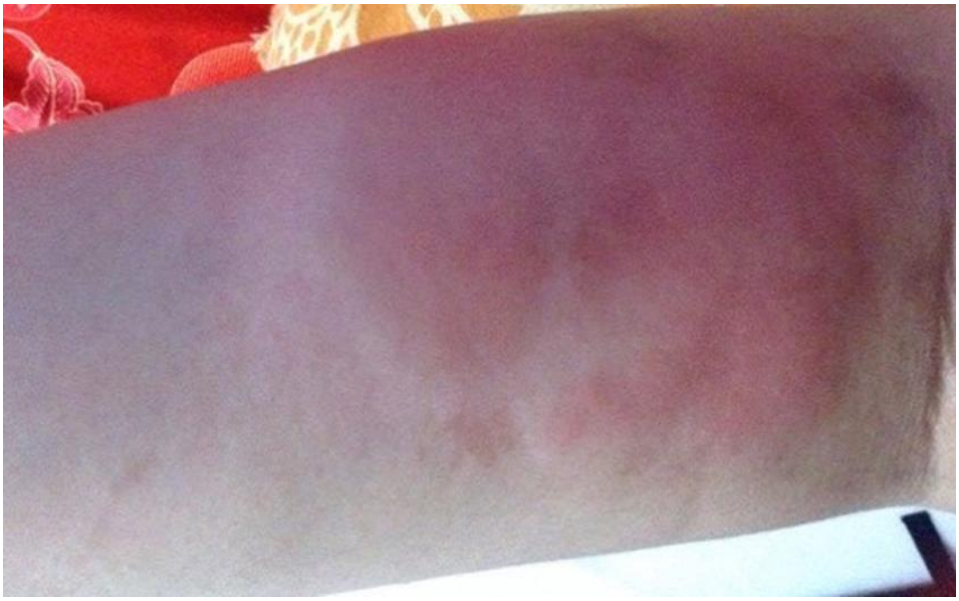

Figure 2.3 A hypersensitivity reaction due to the “bites” of the bed bugs.

2.2 The Next Generation Sequencing Technology

2.2.1 The Emergence of First and Second Generation Sequencing

Deoxyribonucleic acid (DNA) sequencing is a technique that is commonly used to determine the nucleotide sequence of DNA (Prober et al., 1987). This nucleotide sequence is the fundamental information that is important in understanding the genetic function or evolution since it contains the blueprint instructions for building an organism (Sanger, 1977).

The key breakthrough that altered the evolution of DNA sequencing appeared in the mid-1970s (Anderson et al., 1981), which was developed by an English biochemist, Frederick Sanger. This sequencing is the notable first-generation sequencing that utilized dideoxy or Sanger's 'chain-termination' technique (Sanger, 1977). Although traditional Sanger sequencing was the norm for a number of years, numerous initiatives were taken to improve the DNA sequencing method. This is due to Sanger sequencing are limited in throughput and high cost (Metzker, 2010; Morey et al., 2013; Reuter et al., 2015).

The next generation sequencing (NGS) is a high throughput sequencing technology that uses massively parallel sequencing system (Margulies et al., 2005). This sequencing method became available at the onset of 21st century (van Dijk et al., 2014) and yielded a remarkable efficiency, fast and low-priced DNA sequencing platform compared to traditional sequencing technology (Buermans & den Dunnen, 2014; Sucher et al., 2012; Veera Singham et al., 2014). To date, there are various NGS platforms in the market which includes 454 Roche pyrosequencing,

ABI/SOLiD, Solexa/Illumina sequencing (van Dijk et al., 2014; Valouev et al., 2008) and more recently the third-generation sequencing technology, PacBio Sequencing (Eid et al., 2009; Schadt et al., 2010). Of these, 454 Roche pyrosequencing remains the most efficient for the development of microsatellite markers from the genome of non-model organisms, such as bed bugs (Booth et al., 2012; Castoe et al., 2010).

2.2.2 454 Roche Pyrosequencing

454 Roche pyrosequencing platform is the second NGS technology launched after Massively Parallel Signature Sequencing (MPSS) (Barba et al., 2014), and it is the first available in 2004 as a profit-making product (Margulies et al., 2005; Shendure & Ji, 2008). This 454 Roche pyrosequencing method has decreased sequencing cost of 6 times compared to Sanger sequencing methodology. According to Barba et al. (2004), this highly-throughput platform has improved its sequence read from 2005 until 2007. 454 GS 20 Roche sequencing platform was introduced in 2005 – 2006 along with an ability to produce 20 million bases (Mbp) of sequences. This was then replaced in 2007 by announcing GS FLX Roche model, with over 100 Mbp of sequence in only four hours and improved to 400 Mbp in 2008. An upgraded version of 454 GS-FLX+ Titanium was later developed which capable in producing over 600 Mbp of sequence data and Sanger-like read lengths of up to 1000 bp.

Despite the proliferation of sequenced read length due to the technologies advancement (Morozova & Marra, 2008; Barba et al., 2014), the read length of 454 Roche sequencing platform is yet shorter than desirable length when compared to the read length obtained using Sanger sequencing (Luo et al., 2012). Nonetheless, it was mentioned by Gardner et al. (2011) that the main concern in developing genetic

markers are its heterogeneity in marker usability and number of loci. Thus, the read length number produced by 454 Roche pyrosequencing is of less concerned and is reliable for quantitatively evaluating genetic structures within natural communities (Luo et al., 2012).

454 Roche pyrosequencing platform uses pyrosequencing principle as the sequencing chemistry (Gardner et al., 2011; Margulies et al., 2005; Morozova & Marra, 2008; Ronaghi et al., 1998; Shendure & Ji, 2008). As stated in a study by Margulies et al. (2005), pyrosequencing has enables 454 Roche Life Sciences platform to generate approximately five hundred million bases of raw sequence in just a few hours. Pyrosequencing is a sequencing method which relies on the detection of pyrophosphate released during the nucleotides incorporation. The hybridized targeted DNA fragment and the primer are incubated with polymerase, adenosine triphosphate (ATP), firefly luciferase and nucleotide-degrading enzymes (Ronaghi et al., 1998). The process of deoxynucleotide addition occurs repeatedly during incubation. Only a single deoxynucleotide that matches with the base in the template strand will be incorporated into the extending DNA strand and release of pyrophosphate (PPi) is taking place at this point. The PPi then will convert into ATP by ATP sulfurylase (Shendure & Ji, 2008). The concentration of ATP produced is then sensed by enzyme luciferase (Ronaghi et al., 1998). Production of light produced during luciferase-catalyzed reaction can be estimated by a suitable light-sensitive machine. Lastly, the unincorporated deoxynecleotides and ATP produced will subsequently be degraded between each cycle by the nucleotide-degrading enzymes.

2.3 Microsatellite as A Genetic Marker

2.3.1 General Information of Microsatellite

Microsatellite is well-known as simple sequence repeats (SSRs), variable number tandem repeats (VNTRs), short tandem repeats (STRs) (Selkoe & Toonen, 2006) and simple sequence length polymorphisms (SSLPs) which can be found naturally in prokaryotic and eukaryotic genomes (Jeffreys et al., 1985; Tautz & Renz 1984; Thoren et al., 1995; Toth et al., 2000)**.**

This genetic tool has been widely used in many living organisms such as *Escherichia coli* (Gur-Arie et al., 2000; Schlotterer, 2000), humans (Beckmann & Soller, 1990), mice (Love et al., 1990), cows and sheep (Moore et al., 1991), plants (Condit & Hubbell, 1991) and insects (Booth et al., 2012; Hughes & Queller 1993; Veera Singham et al., 2012) for different kinds of basic to applied genetics research. It is biparentally inherited DNA marker which is short, tandemly repeated stretches of DNA of variable, motifs and lengths (Wang et al., 2009). This genetic marker consists of one to six nucleotides in length of tandem arrays (van Oppen et al,. 2000) which is classified into mono-, di-, tri-, tetra-, penta- and hexanucleotide repeats (Abdelkrim et al., 2009). Among different types of tandem arrays, dinucleotide, trinucleotide, and tetranucleotide repeats are most frequently used in molecular genetic studies (Selkoe & Toonen, 2006; Senan et al., 2014). Microsatellites can be further categorized as simple perfect, simple imperfect, compound perfect or compound imperfect (Roy et al., 2004; Senan et al., 2014) repeat motifs.

2.3.2 The Advantages of Microsatellite Marker

Population genetic studies have been remarkably productive and successful in the last decade, and microsatellite marker is one of the most popular and efficient genetic tools in many fields of genetic studies. For use in intraspecific analyses, microsatellite markers have overtaken mitochondrial and other DNA markers currently.

Nowadays, the usage of microsatellite has become more streamlined and less costly. Its improvement in computing technology also enables the researchers to perform an extensive statistical approaches such as maximum likelihood, Bayesian probability theory and Monte Carlo Markov chain simulation (Selkoe & Toonen, 2006). In comparison to the summary statistics of traditional approaches, the usage of microsatellite can use more of the information in a data set due to its mutation rate which range from 10^{-6} to 10^{-2} (Schlötterer, 2000). Thus, the introduction of microsatellites to population genetic studies has advanced the ability to detect population genetic structure and breeding strategy (Booth et al., 2012; Saenz et al., 2012). In addition, they are also able to test parentage and relatedness, to assess genetic diversity and to study recent population history (Zhang & Hewitt, 2000). This crucial knowledge enables us to understand the infestation dynamics, evolution process and diversification of insects which may help us to control pest populations in the future (Wang et al., 2009).

The microsatellites markers show a trait of an ideal markers which is allowing a small tissue sample to be used. Thus, this trait ensure the tissue sample can be easily preserved for future use. As reported by Selkoe and Toonen (2006), DNA-based techniques like microsatellites utilized PCR to amplify the marker of

interest from a minute amount of tissue samples. The stability of DNA compared to enzymes is also an advantageous for microsatellite as a genetic marker (Ariff & Khan, 2009). In addition, the characteristic of having a shorter in length than sequenced loci has resulted in amplifiable PCR despite some DNA degradation (Taberlet & Luikart, 1999). Hence, the usage of microsatellite marker for the samples of ancient DNA or DNA from hair and faecal samples used in non-invasive sampling will be advantageous to the researchers in addressing population genetics (Taberlet & Luikart, 1999). Moreover, microsatellite markers are species-specific in which cross-contamination by non target organisms is likely to occur (Selkoe & Toonen, 2006), compared with universal primers. This trait is beneficial when working with faecal samples or species in which endosymbiont contamination is inevitable (Margulies et al., 2005; Selkoe & Toonen, 2006).

Microsatellite is a single locus, co-dominant markers for which many loci can be efficiently combined in the genotyping process to provide fast and less expensive replicates sampling of the genome (Schuelke, 2000; Weber & May, 1989). This feature can reduce the rate of sampling error since relying on a single locus to evaluate ecological traits from genetic data able to create a high rate of sampling error. Furthermore, taking multiple samples of the genome by combining the results from many loci provides a more precise and statistically compelling technique of comparing populations and inividuals (Bench & Akesson, 2005; Selkoe & Toonen, 2006).

2.3.3 Mutation Model of Microsatellite Marker

Mutation models are important for the analysis of genetic data. It have been proposed that several mutation models of microsatellite can be employed in the analysis of genetic data. According to Kimura and Crow (1964), infinite allelic model (IAM) allows every mutation event to create new allele and the multiple repeats are simultaneously gained or lost. Thus, it assumes that any new allele size within a population has not been encountered previously.

On the other hand, the stepwise mutation model (SMM) assumes that mutations involve the process of adding or subtracting one or more repeat units from the string of repeats at some constant rate. This is called as a one-step symmetric model, with an equal probability of gaining or subtracting. It indicates the process of errors during DNA replication that generates mutations (Ellegren, 2004). This is the usual mechanism of microsatellite mutation.

Besides, non-stepwise mutation processes which include point mutation and recombination events are also can occur (Richard & Paques, 2000). The recombination events can be unequal crossing over and gene conversion. Unequal crossing over is a process where the two chromosome strands are misaligned during crossing over, and it is resulting in a deletion in one DNA molecule and insertion in the other. Such a condition happens when the recombination machinery cannot determine the correct annealing between the two strands (Hancock, 1999). In contrast, gene conversion is a recombination process which occurs when an allele is transferred from one DNA helix (which remains unchanged) to another DNA helix, whose sequence is altered.

2.3.4 The Development of Microsatellite Markers

There are two typical approaches in microsatellites development which are through the construction of a genomic library and developing microsatellites de novo or to search published literature for existing microsatellite primers (Figure 2.5) for the target species (Chambers & McAvoy, 2000; Selkoe & Toonen, 2006; Veera Singham et al., 2014). Constructing a genomic library using enrichment method is a conventional approach that requires time and labor (Abdelkrim et al., 2009). In this enrichment protocol, DNA molecules must be prepared individually for sequencing by cloning and PCR amplification and purification (Senan et al., 2014; Veera Singham et al., 2014). On the other hand, the NGS reduced the time and cost to perform the whole genome sequencing.

According to Margulies et al. (2005) and Gardner et al. (2011), the ability of NGS to generate one million to 43 billion sequence reads is influential in medicine, genetic and evolution study. Due to this, NGS has become a recommended alternative for developing microsatellite over traditional enrichment method (Castoe et al., 2010). NGS is a sequencing platform that does not need a priori information on motif types which is a basic requirement in enrichment technique (Dieringer & Schlotterer, 2003; Gardner et al., 2011). It also skips the use of restriction enzymes that may result in over-representation of fragments containing repetitive elements due to the restriction within transposable elements (Megle´cz et al., 2004; Wilder & Hollocher, 2001). On top of that, it is worth to mention that this second generation sequencing able to retrieve more polymorphic microsatellite loci than the enrichment method (Abdelkrim et al., 2009) and PCR amplification and purification (Senan et al., 2014; Veera Singham et al., 2014).

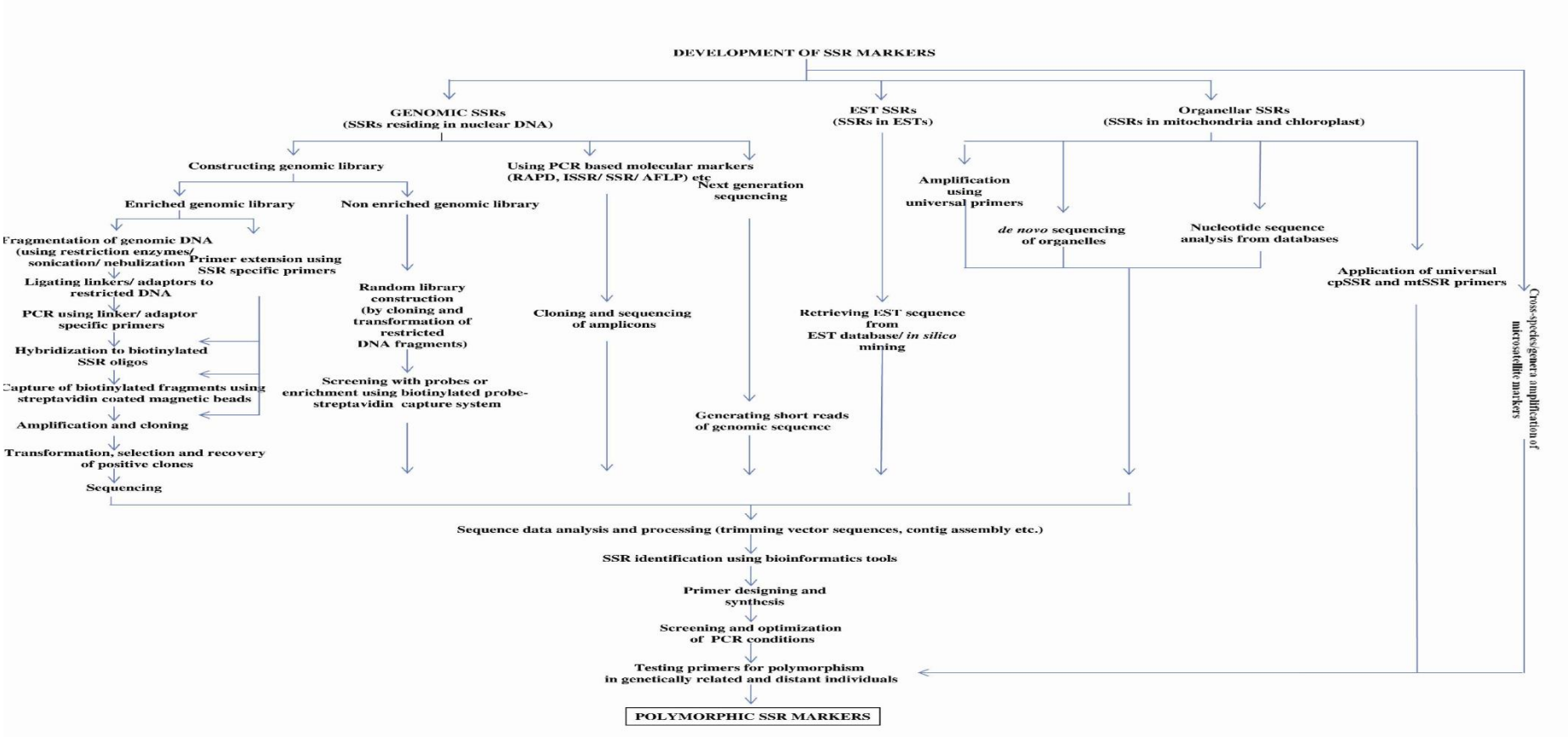

Figure 2.4 A general outline on the development of microsatellite markers (Senan et al., 2014).

Among the various types of NGS technology available, 454 pyrosequencing platform remains the top choice when it comes to the development of novel polymorphic microsatellite markers from the genome of non-model organisms. This is because 454 technology produces longer sequence reads, therefore the resulting raw sequence provide ample size to design primers flanking the microsatellite motifs. Hence, the need for performing sequence assembly as needed in the case of Illumina sequencing which produces short reads 100-150bp can be eliminated and hence save time and cost of the analysis. Additionally, the accuracy of sequence reads from 454 technology is comparable to that of Illumina sequencing as well, therefore reliable in generating robust microsatellite markers for downstream analyses (Hung et al. 2015; Pootakham et al., 2015; Schoebel et al., 2013).

2.4 Mitochondrial DNA as A Genetic Marker

Due to the invention of PCR technology, mtDNA has proven powerful for genealogical and evolutionary studies of animal populations (Freeland, 2005; Stratchan, 1992). Mitochondrial DNA is the power generator of the cell in cell respiration. The oxysomes which located on the cristae of a mitochondria contains ATP synthetase enzyme serve an important role in the process of oxidative phosphorylation (Clary & Wolstenholme, 1985; Tielens et al., 2002). As stated by Rajendra and Akhilesh (2005), the mitochondria DNA comprises of 70% protein, 25-30% lipids, approximately 1% of RNA and less than 1% DNA. A common insect mitochondrial genome consists of 37 genes coding for two ribosomal RNAs, 13 messenger RNAs snd 22 transfer RNAs (Clary & Wolstenholme, 1985) (Figure 2.5).

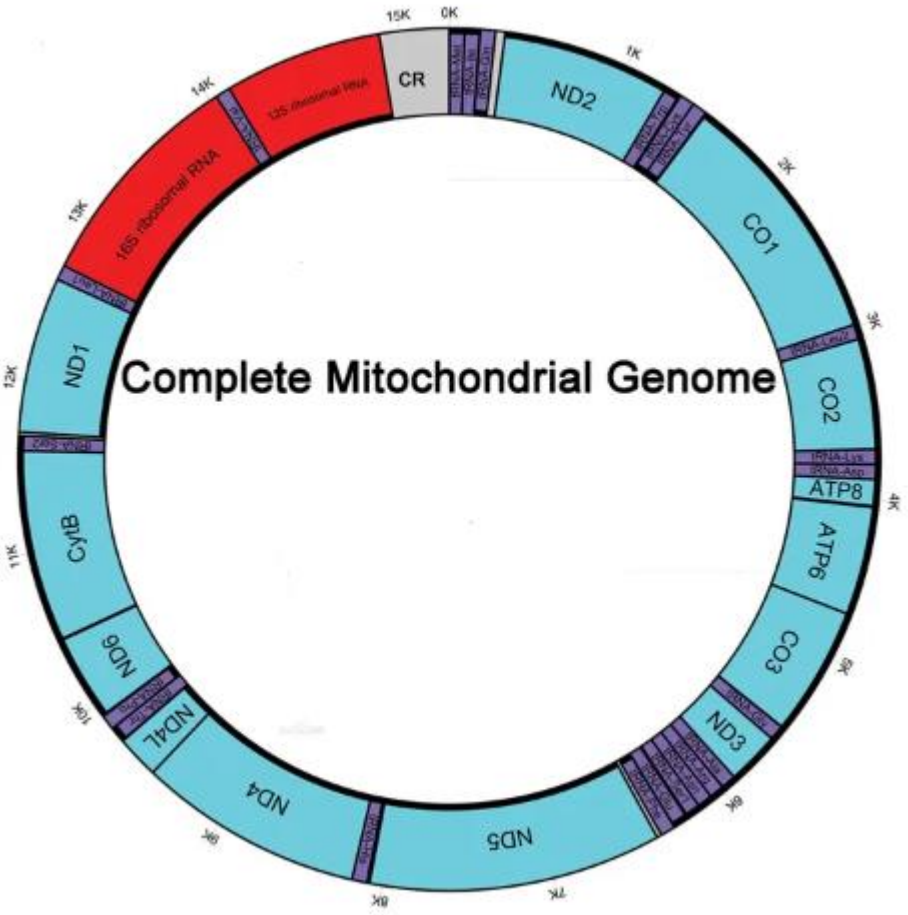

Figure 2.5 A complete mitochondrial genome of insect (Zhao et al., 2017).